The Value of Crowdsourcing: Evidence from Earnings Forecasts

Biljana N. Adebambo and Barbara A. Bliss*

July 2015

Abstract

We use a novel dataset containing earnings forecasts from buy-side analysts, sell-side analysts, and individual investors, to examine whether the crowdsourcing of earnings forecasts provides value-relevant information. Consistent with the 'wisdom-of-crowds' effect, crowdsourced earnings consensus is more accurate than the I/B/E/S consensus 57% of the time. The accuracy of the crowdsourced consensus increases with diversity. The crowdsourced consensus produces errors that are more strongly associated with abnormal returns, suggesting that it is a superior measure of the market's true earnings expectations. A trading strategy based on the difference between the consensuses yields an abnormal return of 0.592% per month.

JEL Classifications: G140, G240

Keywords: Analysts, Earnings Forecasts, Forecast Accuracy, Crowdsourcing, Earnings Announcements

Acknowledgements: We thank Leigh Drogen and the Estimize team for providing us with their data and answering data-related questions. We thank Robert Bowen, Jared Delisle, Jane Jollineau, Sarah Lyon, Johan Perols, Andy Puckett, and Jake Thornock for helpful feedback. All errors are our own.

Biljana N. Adebambo is at the School of Business Administration, University of San Diego, San Diego, CA 92110, Phone: 619-260-4294, Email: bnikolic@sandiego.edu. Barbara A. Bliss is the corresponding author and is at the School of Business Administration, University of San Diego, San Diego, CA 92110, Phone: 619-260-2998, Email: bbliss@sandiego.edu.

The Value of Crowdsourcing: Evidence from Earnings Forecasts

I. Introduction

Quarterly earnings expectations are a key piece of financial information necessary for determining firm value, cost of capital, and expected returns. From a research standpoint, earnings expectations are used as the primary benchmark in event studies to measure the amount of new information that is released at the earnings announcements, and subsequently how efficiently the market incorporates the new information (Kothari, 2001). However, researchers need *accurate* earnings expectations to make correct inferences about market efficiency.

The dominant measure of earnings expectations in finance and accounting literature is the consensus of sell-side analysts' forecasts, despite the well-documented evidence of forecast biases resulting from misaligned incentives and conflicts of interest.[1] The forecast biases are so pervasive that several recent papers find evidence that institutional investors often adjust for these biases when forming their own expectations (Cheng, Liu, and Qian, 2006; Hilary and Hsu, 2013; Malmendier and Shanthikumar, 2007; Mikhail, Walther, and Willis, 2007). These findings suggest that earnings consensus constructed from sell-side analysts' forecasts is not an accurate measure of the market's true earnings expectations.

[1] Literature shows that, among others, sell-side analysts bias their forecasts to gain access to management and information (Chen and Matsumoto, 2006; Ke and Yu, 2006; Lim, 2001;), to benefit the corporate finance side of the investment bank (Chan, Karceski, and Lakonishok, 2007; Dugar and Nathan, 1995; Michaely and Womack, 1999), and for career concerns (DeBondt and Forbes, 1999; Trueman, 1994; Welch, 2000). However, near-term analyst forecasts still appear superior to time series forecasts (Brown, Hagerman, Griffin, and Zmijewski, 1987 and Bradshaw, Drake, Myers, and Myers, 2012)

Crowdsourcing is the process of obtaining services, ideas, or content from a large, undefined group of people rather than from one specific, named group. Recent advances in technology and emergence of social media have facilitated the success of crowdsourcing in many forms such as information production (e.g., Wikipedia), and peer-based opinion generation (e.g., user generated reviews on Yelp.com and Amazon.com). These advances have even encouraged a more prominent role for peer opinions in investment community, which was once dominated by Wall Street professionals (e.g. Seeking Alpha and Stocktwits).

However, studies that examine whether the collective opinions of individual investors convey relevant financial information find mixed evidence. Earlier studies have shown that investor opinions posted on Internet message boards do not meaningfully predict stock returns (Tumarkin and Whitelaw, 2001; Antweiler and Frank, 2004; Das and Chen, 2007). In contrast, Chen, De, Hu, and Hwang (2014) find that the investors' views published on Seeking Alpha predict future stock returns and earnings surprises. Additionally, Gianni, Irvine, and Shu (2014) show that the convergence (divergence) of investors' opinions from Stocktwits posts, is associated with lower (higher) earnings announcement returns.

Similar to Seeking Alpha and Stocktwits, which provide the means to aggregate the opinions of the investment community, Estimize is an online platform that crowdsources earnings and revenue estimates from a wide range of individuals including buy-side analysts from hedge funds, financial institutions, proprietary trading firms, private equity firms, and venture capital firms, sell-side analysts, independent research professionals, and individual investors. Since its inception in 2011, Estimize has gained significant popularity in the investment community, with more than 4,100 of its 34,000 registered users contributing forecasts to the platform.

In this paper, we examine whether the crowdsourcing of earnings forecasts produces value-relevant information. We use the context of the 'wisdom-of-crowds' principle, which suggests that aggregating opinions from diverse, independent, and decentralized sources is likely to produce a more accurate prediction (Surowiecki, 2004), to formalize our analysis.[2] Specifically, we examine three main questions: (1) whether crowdsourced forecasts generate a more accurate consensus; (2) whether the accuracy of the crowdsourced consensus increases with the number and diversity of contributors; (3) whether the crowdsourced consensus is a superior representation of the market's true expectations of earnings.

The 'wisdom-of-crowds' principle suggests that the Estimize (crowdsourced) consensus should be more accurate than the sell-side analyst consensus because it includes forecasts from a broad set of contributors, presumably with diverse and independent opinions. Moreover, given that any registered user can issue an earnings forecast, the earnings consensus from crowdsourced forecasts is likely to capture a portion of expectations that is ignored when solely focusing on the opinions of sell-side analysts. Thus, we expect that this broader, decentralized sample of market participants will produce a more accurate consensus and a more complete representation of the market's earnings expectations, compared to the traditional sell-side consensus.

Alternatively, it is possible that the inclusion of forecasts from certain individuals, such as Non-Professionals, may provide no value, or worse, cause the Estimize consensus to deviate further from actuals. Surowiecki (2004) states that although diversity matters, assembling a group of diverse but thoroughly uninformed people is not likely to lead to wise outcomes. Given the difficulty of forecasting earnings and the information advantage of sell-side analysts, it is unclear

[2] The 'wisdom-of-crowds' principle can be directly observed in several setting, such as information production (e.g., Wikipedia) and election prediction markets (e.g., the Iowa Electronic markets).

whether non-traditional forecasts will contain any unique or superior information that is not already reflected in the traditional analysts' forecasts.[3] Therefore, the value of crowdsourcing earnings forecasts from non-traditional sources remains an open empirical question.

Using a matched sample of firms that have following in both I/B/E/S and Estimize, we find that, on average, the crowdsourced consensus produces smaller absolute forecast errors and is more accurate 57% of the time. Surprisingly, all users, even Non-Professional users, contribute to making the earnings consensus more accurate. Moreover, we show that the consensus accuracy increases with the number of Estimize forecasts and, more importantly, the diversity of contributors, consistent with the 'wisdom-of-crowds' principle.

Second, we find that the crowdsourced consensus better explains the market's reaction to earnings surprises. In the multivariate setting, we find that the Estimize consensus contains significant incremental information about the market's expectations of earnings, especially when the Estimize consensus is comprised of forecasts from diverse contributors. Comparison of the earnings response coefficients (ERCs) shows that the Estimize earnings surprise elicits a 24% stronger market reaction than the I/B/E/S earnings surprise. In situations where the I/B/E/S and Estimize surprise disagree (one is positive and one is negative), the immediate market reaction generates a return of the same sign as the Estimize surprise.

Third, we construct a simple trading strategy based on earnings expectations. Our long-short trading strategy, based on the difference between the Estimize and the I/B/E/S consensus, generates a cumulative abnormal return of 0.592% per month. For the subset of firms that have a diverse following in Estimize, the trading strategy generates a cumulative abnormal return of

[3] Sell-side analysts are likely to have ties with management, expend more effort, and have more financial resources in information gathering.

1.721% per month. Overall, our findings suggest that a broader, diverse group of market participants improves the information set and produces a consensus that is a more accurate representation of the market's true expectations of earnings.

Our paper makes several important contributions. First, we complement recent research on crowdsourcing of financial information. Crowdsourcing is a relatively new phenomenon in the financial industry and the potential benefits are still unknown. On the one hand, research documents that the aggregation of individual investors' opinions and actions can predict future stock returns (Chen, et al. 2014; Hill and Ready-Campbell, 2011). On the other hand, studies have shown minimal correlation between activity on investing platforms and stock performance (Antweiler and Frank, 2004; Das and Chen, 2007; Wang et al., 2014). In contrast to these studies, which examine whether the collective opinions predict future returns or earnings news, we examine whether crowdsourced forecasts improve upon the current measure of earnings expectations from sell-side analysts. This is arguably a higher hurdle because the Estimize earnings consensus must predict future returns and do so better than the traditional consensus. Furthermore, our measure of expectations is numerical and less likely to suffer from any misinterpretation that may occur from using textual analysis to measure investors' opinions. Overall, our paper provides strong evidence on the benefits of crowdsourcing, thereby encouraging the crowdsourcing of a variety of other financial data such as inflation rate, interest rates, GDP, and commodity prices.

Second, we contribute to the literature that examines the quality of sell-side analysts' consensus in comparison to those from other sources. Prior literature has compared sell-side analysts' forecasts to the forecasts from Value Line (Philbrick and Ricks, 1991; Ramnath, Rock and Shane, 2005), independent analysts (Clarke, Khorana, Patel, and Rau, 2008; Cowen, Groysberg, and Healy, 2006; Gu and Xue, 2008; Jacob, Rock, and Weber, 2007); and whisper forecasts (Bagnoli, Beneish, and Watts, 1999; Brown Jr. and Fernando, 2011). The evidence from these studies have been mixed. Bagnoli et al. (1999) and Philbrick and Ricks (1991) are the only two studies to find that alternative sources of forecasts are more accurate than sell-side analysts. However, both of these studies examine the period prior to Reg FD, which affects the generalizability of their findings. It is also worth noting that the above mentioned studies use a much smaller sample of relatively homogenous (e.g. independent analysts) or even unknown contributors (e.g. whispers). Our consensus, on the other hand, contains a wider range of investors' opinions from both, Professionals and Non-Professionals, which was once unobservable. In addition, our sample includes 4,100 forecast contributors following 1,200 firms, which is larger than the samples used in the previous literature. Unlike previous studies that find mixed evidence, we find that the alternative consensus is superior to the traditional sell-side consensus, in both accuracy and in measuring the market's expectations.

Third, we contribute to the finance and accounting literature by introducing a new dataset and a new proxy for earnings expectations that is less affected by sell-side biases and expectations management. Brown (2000) highlights over 575 studies on expectations, most of which are devoted to sell-side analysts' earnings forecasts and stock recommendations. A large number of these studies use the market's immediate response to earnings announcement to examine whether the earnings announcement conveys any new information and how efficiently the market incorporates that information. From an econometrics standpoint, the degree of the return-earnings association is highly dependent upon the proxy of unexpected earnings that researchers employ, and a proxy containing high measurement error is likely to translate into poor explanatory power or lead to erroneous conclusions (Kothari, 2001). Our measure of earnings expectations is likely to have less measurement error and can be measured ex-ante.

A contemporaneous working paper by Jame, Johnston, Markov, and Wolfe (2015) also examines the properties of crowdsourced earnings estimates. However, there are several important differences in our analyses and findings. Unlike Jame et al. (2015), we show that the Estimize consensus alone is more accurate than the I/B/E/S consensus. More importantly, we use the wisdoms of crowds' framework to examine the sources of accuracy, and find that diversity, which is observable ex-ante, is a critical determinant. Although both papers document that the Estimize consensus is a better measure of earnings expectations, we devise a profitable trading strategy to show the full significance of this finding.

The remainder of the paper is organized as follows. Section II outlines prior literature and hypothesis development. Section III describes the data and Section IV presents the empirical results. Section V concludes.

II. Literature Review and Hypothesis Development

The 'wisdom-of-crowds' effect refers to the findings that a large, diverse collection of individuals generally makes predictions better than any single individual, even an expert. The effect is well documented across multiple disciplines with the general notion that the superiority of crowd averages results from the cancelation of idiosyncratic errors (Brown, 1993). However, this effect is contingent upon the properties of the crowd. For example, in the forecasting literature, numerous studies document the benefits of combining forecasts and find that combined forecast embodies the 'wisdom-of-crowds' only if the individual forecasts contain useful and independent information (see surveys by Armstrong, 2001; Clemen, 1989; Timmermann, 2004).

Based on these and similar findings, Surowiecki (2004) formulates four conditions necessary to produce a "wise" crowd: diversity of opinion, independence, decentralization, and aggregation. Diversity implies that each person has their own point of view and some private information, even if only their unique interpretation of the available public information. Diversity is important because it adds different perspectives and increases the amount of available information. Independence requires *relative* freedom from opinions and actions of others, not complete isolation. Independence enables people to actually express their diverse information and reduces potential bias in the group decision. Decentralization allows people to specialize and draw on local knowledge, without any individual or small group dictating the process. Through specialization, decentralization encourages independence and increases the scope and diversity of information. Finally, an aggregation mechanism is necessary to collect the individual opinions and harness the 'wisdom-of-crowds' effect.

Fortunately, the Estimize platform enables all four elements of crowd wisdom to exist in the process of setting the earnings consensus. Specifically, by allowing any individual to contribute their estimate of earnings, Estimize promotes decentralization and diversity of opinion. Indeed, biographical data of Estimize contributors indicate that contributors come from various institutions and professional backgrounds (see section III.B). Further, the freedom to cover any firm allows the decentralized Estimize contributors to draw upon any expertise, special local/industry knowledge, or interest that they may have when forming an estimate. This decentralization promotes independence. In addition, the fact that Estimize users' compensation and career outcomes are not directly tied to their earnings forecasts on Estimize, should make them less likely to be influenced by other's opinions.[4]

[4] Consistent with this idea, most of the contributors that we contacted feel that there are no significant costs associated with contributing on Estimize and they are primarily motivated by desire to beat Wall Street and their peers, which

In contrast to Estimize, traditional sell-side analyst are unlikely to exhibit the properties of a wise crowd due to incentives and conflict of interests. Empirical evidence show that analysts tend to herd by releasing forecasts similar to those previously announced by other analysts (DeBondt and Forbes, 1999; Trueman, 1994; Welch, 2000), reducing independence. In addition, analysts may strategically bias their information to improve management relations, among other reasons (Chen and Matsumoto, 2006; Cotter, Tuna, and Wysocki, 2006; Das, Levine, and Sivaramakrishnan, 1998; Francis and Philbrick, 1993; Ke and Yu, 2006; Lim, 2001; Matsumoto, 2002). Further, the diversity of opinion is likely to be limited since sell-side analysts are likely to draw upon the same information resources and use similar models. Consequently, the aggregation of forecasts across a wide range of contributors through a platform like Estimize is likely to alleviate the above mentioned issues associated with sell-side analysts' forecasts. This leads us to our first hypothesis:

H1: Crowdsourcing improves upon the forecast accuracy of earnings consensus, beyond the accuracy of sell-side analysts.

The benefits of combining individual forecasts are highly dependent upon the number of independent forecasts and the additional information contained in each forecast (Armstrong, 2001). For example, Batchelor and Dua (1995) found that the accuracy of macroeconomic variable forecasts increased 9% when combining any two economists' forecasts, and by 16.4% when combining ten individual economists' forecasts. In addition, when Batchelor and Dua (1995) combined the forecasts of economists with different backgrounds, the reduction in forecast error

suggests that their estimates should be bold and independent. Additionally, Jame et al. (2015) find that the individual Estimize forecasts tend to be bolder.

was greater than when they combined the forecasts of economists with similar backgrounds. These findings suggest that amount of additional information contained in each forecast is a function of each contributor's background. Therefore, the consensus forecast error is likely to decrease as the diversity and number of the forecasts increases. This leads us to our second hypothesis:

H2: Increases in the number and diversity of forecast contributors will increase the accuracy of earnings consensus.

Investors form their expectations by weighing different sell-side analysts' forecasts based on the perceived quality. Given that some corporate managers often guide analysts' earnings forecasts downward to avoid missing earning expectations (Cotter et al., 2006; Matsumoto, 2002; Richardson, Teoh, and Wysocki, 2004), market participants may not rely as much on these forecasts when forming earnings expectations. Indeed, recent literature finds that institutional investors often adjust for these biases when forming their own earnings expectations (Cheng et al., 2006; Hilary and Hsu, 2013; Malmendier and Shanthikumar, 2007; Mikhail et al., 2007). These findings indicate that the analyst consensus may not adequately represent the expectations of the largest and most active segment of investors. Estimize contributors, on the other hand, do not face similar bias-inducing incentives. Additionally, Estimize contributors are likely to represent a broader segment of the market because any individual can contribute their forecast. This leads to our third hypothesis:

H3: Crowdsourced earnings consensus is a better measure of the market's true expectations of earnings.

III. Data and Sample Construction

A. Estimize Institutional Details and Data

Estimize is an open online platform that crowdsources quarterly earnings and revenue estimates from a wide range of contributors. Estimize started in late 2011 by populating their platform with 2,700 stocks and inviting buy-side analysts and portfolio managers to contribute their forecasts for any of those stocks. In a short time, Estimize has gained significant popularity in the investment community with over 34,000 registered users, more than 4,100 of whom have contributed at least one earnings forecast on the platform.[5] Besides being available directly on their website, Estimize earnings and revenue consensus estimates are uploaded onto Bloomberg terminals and often reported alongside with the Wall Street consensus in news outlets. For example, a recent Yahoo! Finance news article on Netflix's upcoming earnings announcement reported that "The Estimize community forecasts earnings per share (EPS) of $0.46 compared to the Wall Street consensus of $0.32. In terms of revenue, Estimize predicts a figure of $1.653 billion, slightly above the Wall Street number of $1.646 billion."[6]

To illustrate the Estimize platform, Figure 1 presents example data for Lululemon Athletics Inc. Estimize users are able to view the upcoming earnings announcement date, the past quarterly earnings of the company, the company's guidance (if provided) for current and past quarters, the Wall Street consensus for current and past quarters, and the Estimize consensus for current and past quarters.[7] In addition, Estimize users can see how many forecasts are included in the Estimize

[5] Estimize has been featured on CNBC, the Wall Street Journal, CNN Money, Forbes, the Economist, Fortune, Businessweek, Barron's, and the CFA Institute newsletter.

[6] http://finance.yahoo.com/news/netflix--intel-could-surprise-wall-street-181040001.html#

[7] Wall Street Consensus on Estimize is obtained from Zack's.

consensus and view the individual forecasts of all contributors. Any registered user is able to contribute earnings and revenue forecasts on the Estimize platform for any number of firms and at any frequency they choose.[8]

[Insert Figure 1]

This flexibility and openness of the platform, however, could have some disadvantages. Specifically, if anyone, including retail investors and students, is allowed to contribute their earnings forecasts, the quality of these forecasts may be inferior to those issued by professional sell-side analysts. Estimize users are pseudo-anonymous, which makes it difficult to determine the users' information sets and forecasting skills. Hence, it is unclear whether the average Estimize user has superior information or forecasting ability that would make this source of information valuable to investors or researchers. In addition, one may wonder why an individual would be willing to share their superior information with the Estimize community.

We believe that there are several possible incentives that could explain willingness to contribute accurate forecasts. First, there is a shared understanding among contributors that if they contribute, others will as well. Therefore, many contribute their forecasts to be able to obtain the forecasts of the other contributors.[9] The second possible incentive to contribute is reputation building. Estimize is a way for many contributors to create a verifiable track-record of their forecasting ability and gain exposure among their peers.[10] Finally, competitiveness and desire to

[8] If a contributor wishes to issue an earnings forecast for a company that is currently not covered on the Estimize platform, they can contact Estimize and the company will be added to the platform.

[9] Estimize sends to contributors the consensus and forecast updates for the companies they contribute for.

[10] Estimize platform has a visible accuracy ranking of the contributors based on all forecasts made, which includes user summary statistics of error rate, accuracy percentile, and the number of estimates. In addition, Estimize will sometimes feature accurate contributors on podcasts.

voice opinions and correct others' misconceptions provide motivation for some contributors. Estimize is structured as a game with the ultimate goal of being more accurate than the Wall Street consensus.

To gather anecdotal evidence on what motivates users to contribute accurate forecasts, we asked 30 random Estimize users with a track record of being accurate why they contribute to the Estimize platform.[11] Out of 30 requests, 8 "Professional" and 2 "Non-Professional" users responded. Most users provided multiple reasons for contributing. 70% of respondents stated competition as motivation, 50% stated that they use Estimize to build a verifiable track record, 50% stated the desire to improve the earnings consensus, and 20% stated the "fun" element.

It is also possible that some contributors may contribute with the desire to game the system and manipulate investors' opinions of corporate earnings.[12] Although we cannot completely rule out this incentive, Estimize has several quality checks in place to ensure accuracy and to prevent such erroneous forecasts from entering the dataset. Specifically, Estimize uses several algorithms to detect and prevent any suspicious activity such as collusion (clustering of forecasts), the creation of multiple accounts, outlier estimates based on the history of earnings surprises, and estimates generated from bots. These quality checks should mitigate concerns about data integrity. To the

[11] We are interested in why accurate individual forecasters are potentially willing to give up their information advantage. Hence, we selected random individuals from the Estimize "Rankings" page who had Linkedin accounts connected to their Estimize profiles. We asked the open ended question: "Why do you contribute earnings forecasts on Estimize (what motivates you to contribute)?

[12] For example, a short-seller may contribute a low earnings forecast to cause a drop in stock price and profit from his short position. Alternatively, a fund manager holding a stock may contribute a high estimate to boost price.

extent that some gaming influences may still exist in the data, they would bias against finding results.

The Estimize dataset contains a unique identifier for each forecast, contributor, and earnings' release event. For each forecast provided by users, the dataset contains the forecasted earnings per share, the date and time the estimate was issued, the fiscal year and quarter of the earnings announcement, the earnings announcement date, and the official ticker symbol of the firm. The Estimize dataset also contains biographical data for the users who wish to identify themselves as a "professional" or "non-professional" user, however the names of the institutions pertaining to the users are not disclosed. Professional users, who are validated through their work email accounts, can identify their area of work, such as Hedge Fund, Mutual Fund, or Independent, and Non-Professional users can select their sector background, such as Information Technology, Consumer Staples, or Telecommunications. Only about 5.05% of the estimates are generated by contributors that do not provide any biographical information, and we group those users with Non-Professional users. Our Estimize sample includes 57,855 earnings forecasts for 7,528 firm-quarter observations from 4,131 unique contributors.[13]

Given the recent emergence of the platform, we begin by examining the trends of coverage in Estimize. Figure 2a displays a number of Estimize contributors over time. The figure shows a significant increase in the number of contributors from 83 in Q1:2012 to around 800 in Q4:2014. Approximately 64% of the contributors are Non-Professionals and 36% are Professionals. An average contributor issues 9 forecasts per quarter, although the number varies significantly over time and by investor type, as shown in Figure 2b. For example, Professionals have become more

[13] These numbers include forecasts made within 14-days of an earnings announcement for firms that have I/B/E/S and CRSP coverage, and satisfy appropriate filters described in Section III.B.

active over time with the average number of forecasts increasing from 4.78 in Q1:2012 to 15.92 in Q4:2014. Non-Professional contributors, on the other hand, increase their activity over the first part of the sample and then scale back over the second part of the sample, for approximately the same number of estimates per contributor at the beginning and the end of the sample period. Increases in the number of contributors and their activity have led to an increased breadth of coverage. As Figure 2c shows, initially only about 260 firms attracted crowd coverage with the number increasing to 1,200 firms by the end of our sample period.[14] Finally, Figure 2d shows a trend in the average number of forecasts per firm by contributor type. The average number of forecasts per firm has tripled from approximately 4 in Q1:2012 to over 13 in Q4:2014. Overall, evidence from Figure 2 suggests sizeable depth, breadth, and diversity of coverage in Estimize.

[Insert Figure 2]

B. Sample and Variable Construction

We begin our sample construction by obtaining one-quarter ahead earnings forecasts, actual earnings, and announcement dates from the I/B/E/S unadjusted detail and actual files.[15] Next, we obtain stock price, volume, shares outstanding, share code, industry code, ticker symbol, and cumulative adjustment factor data from the Center for Research in Security Prices (CRSP). Finally, we merge Estimize forecasts by ticker symbols and manually confirm the validity of the

[14] Our sample stops in October 2014, so the coverage information for the fourth quarter of 2014 is incomplete.

[15] We merge I/B/E/S information with quarterly financial-statement data from Compustat, and following Dellavigna and Pollet (2009) set the earnings announcement date to the earlier of the announcement dates reported in Compustat and I/B/E/S.

ticker merging. We require each firm in our sample to have I/B/E/S and CRSP coverage, and restrict our sample to common stocks (share codes 10 or 11) with a share price greater than $1.

To prevent the influence of stale forecasts, we only keep Estimize and I/B/E/S forecasts issued within 14 days prior to the earnings announcement. If an I/B/E/S analyst or an Estimize contributor issues multiple forecasts for a given firm-quarter, we only keep the most recent forecast issued. To prevent data errors, we eliminate observations where the actual earnings or forecasts are greater than the stock price and remove observations where the actual earnings reported in I/B/E/S and Estimize differ by more than one cent. The full I/B/E/S sample consists of 27,905 unique firm-quarter observations during 2012-2014. The sample period is determined by the availability of Estimize data. Of the 27,905 firm-quarter observations, 7,528 firm-quarter observations have forecast contributions on Estimize.

To examine the influence of diversity, we construct a measure that utilizes the background of the Estimize contributors. *Diversity* is the number of unique backgrounds of contributors whose forecasts are included in the consensus. For Estimize, *Diversity* can range from 1 to 29, encompassing the following biographical backgrounds provided by the users: asset manager, broker, endowment fund, financial advisor, fund of funds, hedge fund, independent research firm, insurance firm, investment bank, mutual fund, pension fund, private equity, proprietary trading firm, venture capital, wealth manager and other for professionals; academia, consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, student, telecommunication services, utilities, and non-disclosed for non-professionals. For I/B/E/S, diversity is equal to one since all analysts are sell-side. For Estimize, each background of the contributor is only counted once. For example, if the Estimize consensus only contains two forecasts by separate Hedge Fund buy-side analysts, the diversity measure is equal to 1. If the Estimize consensus contains a forecast from a Hedge Fund buy-side analyst and an individual investor from the Telecommunications sector, the diversity measure is equal to 2.

C. Descriptive Statistics

Panel A of Table 1 reports the firm characteristics for I/B/E/S firms (the full universe; Column 1), I/B/E/S firms with Estimize following (Column 2), and I/B/E/S firms without Estimize following (Column 3).[16] Column 4 (5) presents test statistics for difference in mean (median) characteristics between I/B/E/S firms with Estimize following and those without Estimize following. I/B/E/S firms with Estimize following are larger, have more I/B/E/S coverage, and more accurate earnings forecasts from I/B/E/S analysts than I/B/E/S firms without Estimize following. These characteristics capture the availability of information and ease of forecasting, and suggest that Estimize contributors follow stocks with better information environments. In addition, I/B/E/S firms with Estimize following are more growth oriented and trade more frequently than those without Estimize following, as demonstrated by book-to-market ratio and share turnover. This finding shows that the Estimize contributors actively follow firms that are associated with sell-side biases such as growth firms (Chan et al., 2007) and firms that may generate higher trading commissions (Jackson, 2005). Moreover, I/B/E/S firms with Estimize following have much higher (more positive) signed forecast errors and lower forecast dispersion, providing further evidence that Estimize contributors may choose to follow firms that are more likely to suffer from sell-side analysts' bias and herding.

[Insert Table 1]

[16] We do not report characteristics for firms that have Estimize coverage but do not have I/B/E/S coverage because the number of observations with non-missing characteristics is small and we do not use those observations in any tests.

Panel B of Table 1 contains the forecast summary statistics for the matched sample of firms that have I/B/E/S and Estimize following within 14-days of the announcement. The average number of I/B/E/S forecasts is 6.00, which is similar to the number of Estimize forecasts (5.59). The average (median) diversity measure for the Estimize forecasts made within 14-days of the announcement is 3.52 (4). In unreported results, we find that the Estimize forecasts are usually issued closer the announcement date than the I/B/E/S forecasts. For example, 75.77% of all Estimize forecasts are issued within 14-days of the announcement, whereas 46.07% of all I/B/E/S forecasts are issued within 14-days of the announcement. These distribution differences highlight the Estimize contributors' flexibility and ability to include new information into their estimates.

IV. Empirical Results

A. Crowdsourcing and Forecast Accuracy

Our first hypothesis posits that a broader population of contributors should predict future earnings more accurately than a narrower and more homogenous population of the sell-side analysts that are captured in I/B/E/S. To examine this hypothesis, we compare the accuracy of the Estimize contributors' consensus to the accuracy of the I/B/E/S sell-side analysts' consensus for a paired sample of firms that have both Estimize and I/B/E/S following. In addition, we take advantage of the Estimize users' biographical data to examine whether accuracy is concentrated among a select subset of contributors.

We measure accuracy using absolute forecast error, which is defined as the absolute difference between the actual announced earnings obtained from I/B/E/S and the earnings consensus, normalized by the share price at the end of the corresponding quarter (Kothari, 2001).[17] We construct the earnings consensus based on the mean forecast issued within 14-days prior to the announcement and calculate the forecast errors separately for the I/B/E/S consensus and the Estimize consensus.[18]

We perform several univariate tests, which are reported in Table 2. Panel A of Table 2 shows the mean (median) absolute forecast errors for I/B/E/S and Estimize. The average (median) absolute forecast error for the I/B/E/S consensus, $AFE^{I/B/E/S}$, is 0.190 (0.083). The average (median) absolute forecast error for the Estimize consensus, $AFE^{Estimize}$, is 0.186 (0.077). To test the statistical difference of absolute forecast errors between I/B/E/S and Estimize, we use a paired t-test for means and Kruskal-Wallis test for medians. Both tests show that the differences in absolute forecast errors are significantly positive, indicating that the I/B/E/S consensus produces significantly larger errors than the Estimize consensus. Given the average stock price of $52.191 (see Table 1), this difference in average absolute forecast errors translates into a difference of 0.21 cents or 0.30% of the average actual earnings of $0.707.

[Insert Table 2]

To identify the source of accuracy in Estimize, we construct two separate Estimize consensuses by user type. We construct the $Estimize^{Professional}$ consensus based on the forecasts issued by Professional contributors and the $Estimize^{Non\text{-}Professional}$ consensus based on the forecasts issued by Non-Professional contributors. The average (median) absolute forecast error from the Professional consensus, $AFE^{Estimize\text{-}Professional}$, is 0.167 (0.068); and the average (median) absolute

[17] Actual earnings, earnings forecasts, and stock prices are per share values adjusted for splits using CRSP cumulative adjustment factor.

[18] In the robustness tests section, we show that our results are similar if we use median forecast for consensus.

forecast error from the Non-Professional consensus, $AFE^{Estimize\text{-}Non\text{-}Professional}$, is 0.184 (0.077). The Non-Professional consensus is lower than the full Estimize consensus and the I/B/E/S consensus, but is larger than the Professional consensus. This result is possibly driven by the following of different firms.

To examine this possibility, we restrict our sample to the 5,016 firm quarter observations that are covered by Professional users and test the difference in absolute forecast errors. Panel B of Table 2 reports these results. Surprisingly, when we combine the Non-Professionals' and the Professionals' forecasts, the absolute forecast errors decrease by 0.002 (*p*-value=0.02), from 0.167 for Professionals only to 0.165 for the overall Estimize consensus. This result suggest that all users, including Non-Professional users, contribute to making the consensus more accurate, demonstrating the 'wisdom-of-crowds' effect.

Our second hypothesis states that forecast accuracy should increase with the number and diversity of contributors. To test this hypothesis in the univariate setting, we examine forecast accuracy by *Diversity*, a number of unique backgrounds of Estimize contributors. Specifically, each quarter we sort firms into terciles based on *Diversity* and report the average (median) absolute forecast error for the subsamples of firms that have high (top tercile) and low (bottom tercile) diversity. Panel C of Table 2 displays the results. For firms with highly diverse contributors, the average absolute forecast error is 0.133, which is much lower than the absolute forecast error for firms that have a less diverse following (0.232). More importantly, we find that the absolute forecast error is significantly lower for the Estimize consensus than for the I/B/E/S consensus in the high diversity sample, but significantly greater for the Estimize consensus than for the I/B/E/S consensus in the low diversity sample. The difference in the absolute forecast errors for the high-diversity sample translates into 0.8 cents on average or 1.13% of average actual earnings. These results provide initial support that diversity helps the consensus converge to the correct answer.

To ensure that increases in accuracy is not driven by large improvements for a few observations, we report the frequency of the differences in absolute forecast errors between I/B/E/S and Estimize, and the associated binomial test for difference in proportions in Panel D of Table 2. A positive difference indicates that the Estimize consensus is more accurate than the I/B/E/S consensus. For the full sample, we find that the difference is positive for 4,137 observations and negative for 3,157 observations. The earnings consensus constructed using Estimize forecasts is more accurate than the traditional I/B/E/S consensus 57% of the time, which is significantly different from 50% (p-value=0.00). For firms with professional coverage, the Estimize consensus is more accurate than the I/B/E/S consensus 58% of the time. Finally, for firms with a diverse following, the Estimize consensus is significantly more accurate than the I/B/E/S consensus 61% of the time. These findings suggest that Estimize consensus is more accurate in the majority of cases, which provides initial validation of its usefulness for investment and research applications.

B. Impact of the Number and Diversity of Contributors on Forecast Accuracy – Multivariate Analysis

Our second hypothesis states that the number of estimates used in the consensus should affect the accuracy of the consensus. In addition, holding the number of estimates constant, greater diversity of the contributors who submit forecasts should lead to increased accuracy. In this section, we use a multivariate regression analysis on the paired sample of firms that have both I/B/E/S and Estimize following to examine these predictions in more detail. Our dependent variable is the absolute forecast error for the I/B/E/S or Estimize consensus. Specifically, we estimate the following pooled Tobit regression equations:

$$Consensus\ AFE = \beta_0 + \beta_1(E) + \beta_2(\#Analysts^{Estimize}) + \beta_3(\#Analysts^{Estimize} \times E) + \beta_4(\#Analysts^{I/B/E/S}) + \beta_5(Horizon) + \beta_6(\#Analysts^{I/B/E/S} \times E) + \beta_7(Horizon \times E) + Calendar\ Fixed\ Effects + \varepsilon \quad (1)$$

$$Consensus\ AFE = \beta_0 + \beta_1(E) + \beta_2(Diversity) + \beta_3(\#Analysts^{I/B/E/S}) + \beta_4(Horizon) + \beta_5(\#Analysts^{I/B/E/S} \times E) + \beta_6(Horizon \times E) + Calendar\ Fixed\ Effects + \varepsilon \quad (2)$$

$$Consensus\ AFE = \beta_0 + \beta_1(E) + \beta_2(\#Analysts^{Estimize}) + \beta_3(\#Analysts^{Estimize} \times E) + \beta_4(Diversity^{\perp}) + \beta_5(\#Analysts^{I/B/E/S}) + \beta_6(Horizon) + \beta_7(\#Analysts^{I/B/E/S} \times E) + \beta_8(Horizon \times E) + Calendar\ Fixed\ Effects + \varepsilon \quad (3)$$

In the first equation, which tests our first prediction, our main variables of interest are E and the interaction term $\#Analysts^{Estimize} \times E$. E is a binary variable equal to one if the absolute forecast error is from the Estimize consensus, and equal to zero otherwise. $\#Analysts^{Estimize}$ is the number of Estimize contributors who have issued an earnings forecast for a particular firm-quarter within 14 days of the earnings announcement. We expect the coefficients on these two variables to be negative and significant.

For the second prediction, our main variable of interest is *Diversity*, the number of unique backgrounds for contributors whose forecasts are included in the consensus. We expect the coefficient on *Diversity* to be negative and significant. *Diversity* is highly correlated with the $\#Analysts^{Estimize}$ hence, in our full specification (equation 3), we use residual diversity ($Diversity^{\perp}$). $Diversity^{\perp}$ is a residual from the regression of Diversity on E, $\#Analysts^{Estimize}$, and $\#Analysts^{Estimize} \times E$.

In all specifications, we control for the number of I/B/E/S forecasts ($\#Analysts^{I/B/E/S}$), median number of days between forecast issuance and earnings announcement (*Horizon*), and the interactions between Estimize indicator and the above controls. We include $\#Analysts^{I/B/E/S}$ and interact it with the *Estimize* indicator because competition among analysts should promote accuracy of the I/B/E/S forecasts. We control for *Horizon* because accuracy should increase with a proximity to the announcement date (Cooper, Day, and Lewis, 2001; Lehavy, Li, and Merkley, 2011; Richardson et al., 2004). We interact *Horizon* and the Estimize indicator to control for any difference in forecast horizons between the Estimize and I/B/E/S analysts. We also include calendar fixed effects and cluster standard errors by the announcement date.[19]

[Insert Table 3]

Table 3 presents the results.[20] As expected, $\#Analysts^{I/B/E/S}$ is negatively associated with absolute forecast errors in all specifications, and more so for absolute forecast errors from the I/B/E/S contributors. *Horizon* and the interaction term between *Horizon* and *Estimize* are statistically significant and suggest that the Estimize contributors issue more accurate forecasts closer to the earnings announcement, possibly due to the incorporation of relevant new information.

More importantly, we find that the coefficient for E is negative and significant at the 1% or 10% level in all specifications, indicating that the Estimize consensus is more accurate overall. Furthermore, in the first model (column 1), the coefficient on the interaction term $\#Analysts^{Estimize} \times E$ is negative and significant at the 5% level, suggesting that as more contributors participate in the information gathering process of earnings estimates, the consensus becomes more accurate.[21] To gauge the economic significance of this finding, consider that the standard deviation

---

[19] We use a paired sample of firms, which reduces the need to control for firm-specific determinants of forecast accuracy such as size, book-to-market, profitability, institutional ownership, and cash-flow volatility.

[20] Results are similar if we use the natural log of $\#Analysts^{I/B/E/S}$ and $\#Analysts^{Estimize}$.

[21] Results are similar if we use the natural log of #Analysts variables instead.

of the number of Estimize contributors for the 14-day horizon is 7.22, and the average absolute forecast error for Estimize contributors in the 14-day horizon is 18.6.[22] The coefficient on *#Analysts*^{Estimize} of -0.247 and the interaction term of -0.059 then indicate that a one-standard-deviation increase in the number of contributors in Estimize decreases the absolute forecast error to 16.39. This decrease represents a 11.88% drop in absolute forecast errors and translates into 1.15 cents or 1.63% of average earnings.

In the second model (column 2), the coefficient on *Diversity* is -0.993 and significant at the 1% level. The coefficient suggests that one-standard-deviation increase in the *Diversity* of Estimize contributors reduces the absolute forecast error by 2.91 or 15.64%. This reduction represents 2.15% of average earnings. Moreover, the coefficient on *Diversity*$^{\perp}$ in column 3 is also highly economically and statistically significant at -1.716 (*p*-value=0.00), suggesting that greater diversity further improves accuracy, when holding the number of forecasts constant. Overall, the results in Table 3 are consistent with our second hypothesis, and suggest that researchers and investors will benefit from the Estimize consensus, especially when it includes more and diverse forecasts.

C. Why Are Crowds Wiser?

In Tables 2 and 3, we show that the Estimize consensus is more accurate than the I/B/E/S consensus. In this section, we examine a possible explanation for this improvement in accuracy. Literature shows that I/B/E/S analysts tend to be overly pessimistic to allow the firms to easily beat their forecasts (Matsumoto, 2002; Richardson et al., 2004). We argue that Estimize contributors do not have similar incentives and that their consensus is more accurate, at least in

[22] The 14-day error is multiplied by 100 to be at the same scale as the coefficients.

part, because it does not include the same bias. To examine this explanation, we analyze the proportion of positive and negative forecast errors separately for observations where the Estimize consensus is more accurate and for observations where the I/B/E/S consensus is more accurate. Figure 3 depicts the results.

[Insert Figure 3]

The first portion of the figure shows the observations where the Estimize consensus is more accurate ($AFE^{Estimize}<AFE^{I/B/E/S}$). We find that both, the Estimize and I/B/E/S consensus, tend to be pessimistic more often than optimistic. However, the Estimize consensus is pessimistic in only 64.74% of cases while the I/B/E/S consensus is pessimistic in 86.06% of cases. The difference in the proportion of pessimistic errors is economically very significant and unreported χ^2 test shows that it is statistically significant at the 1% level. In contrast, the chart for the observations for which the I/B/E/S consensus is more accurate shows that both Estimize and I/B/E/S tend to be optimistic more often than pessimistic. The proportion of pessimistic forecasts is still higher for I/B/E/S at 42.59% than for Estimize at 35.14%, but the two proportions are much closer together. Overall, Figure 3 shows that the Estimize consensus is less pessimistic than I/B/E/S consensus, and that the I/B/E/S consensus is less accurate when it is more pessimistic, suggesting that the accuracy of the Estimize consensus is in part driven by the correction of the inherent biases in the I/B/E/S consensus.

D. Crowdsourced Consensus as a Superior Measure of Market Earnings Expectations

Our last hypothesis proposes that the crowdsourced consensus is a superior measure of the market's true expectations of earnings than the I/B/E/S consensus because it comes from a broader and more diverse set of market participants. To test this hypothesis, we examine the immediate market reactions to the earnings surprises from I/B/E/S and Estimize. We measure the immediate market reaction using cumulative abnormal returns (CAR). CAR[0,1] is the difference between the buy-and-hold return of the announcing firm and that of a size and book-to-market (B/M) matching portfolio over the window [0,1] in the days surrounding the announcement date.[23]

We begin by examining whether the Estimize consensus contains incremental information about market earnings expectations. Following prior literature, we use decile rank of earnings surprise. Specifically, each quarter we independently sort firms into deciles based on the I/B/E/S earnings surprise (FE_{10}) and the Estimize earnings surprise (FE_{10}^{E}). For comparison with prior studies, our surprise variable is FE_{10}. We use a difference in earnings surprise deciles between Estimize and I/B/E/S (FE_{10}^{E}-FE_{10}), which we label as FE_{10} *Difference*, to examine whether the Estimize consensus contains incremental information about the market's expectations. Specifically, we estimate the following regression:

$$CAR[0,1] = \beta_0 + \beta_1(FE_{10}) + \beta_2(FE_{10}\ Difference) + \sum_{i=1}^{n} a_i Control_i + \sum_{i=1}^{n} c_i(FE_{10} \times Control_i) + Industry + Month + Year + \varepsilon \quad (4)$$

Following DellaVigna and Pollet (2009) and Drake, Gee, and Thornock (2014), our set of control variables includes decile of the firm size, decile of the firm's book-to-market ratio, I/B/E/S analyst coverage, share turnover, reporting lag, indicator for Friday announcements, interaction of each of these variables with FE_{10}, industry fixed effects based Fama-French 10-industry classification, and month and year fixed effects. A detailed description of the variables is located in the Appendix. Our main variable is FE_{10} *Difference*. The difference is positive (negative) when the earnings surprise decile based on the Estimize consensus is higher (lower) than the earnings

[23] Following DellaVigna and Pollet (2009), we exclude observations with returns in the top and bottom 5/10,000th of the distribution for the window. Our results are similar if we winsorize instead.

surprise decile based on the I/B/E/S consensus, indicating a more positive (negative) surprise. If the I/B/E/S consensus perfectly captures the market's expectations of earnings, then the difference should be insignificant. In contrast, if the Estimize consensus contains incremental information about the market's expectations of earnings, then we expect the coefficient on FE_{10} *Difference* to be positive and significant.

Table 4 reports the pooled OLS regression results. In Column 1, we estimate a benchmark regression without FE_{10} *Difference*. As expected, we find a strong positive relation between the announcement-day abnormal cumulative returns and the I/B/E/S earnings surprise decile (FE_{10}). The decile of firms with the most positive earnings surprise outperforms the decile of firms with the most negative earnings surprise by 14.4% on average. More importantly, in Column 2, we find that the coefficient on FE_{10} *Difference* is also positive at 0.005 and statistically significant at the 1% level. This coefficient suggests that firms, which are ranked one decile higher (lower) by the Estimize earnings surprise than by the I/B/E/S earnings surprise, will experience 0.5% higher (lower) returns in the 2-day window around the earnings announcement. In an extreme case, a firm in the most positive decile for I/B/E/S and the most negative decile for Estimize will realize 4.5% *lower* return compared to the firms that are in the most positive decile for both earnings surprises.

[Insert Table 4]

We find additional evidence on the incremental power of the Estimize consensus to explain the market's reactions to earnings surprises when we examine the R^2 of the two regressions. FE_{10} and controls explain 12.7% of the variation in the announcement-day abnormal returns. Adding FE_{10} *Difference* increases the R^2 to 14.2%. The increase in R^2 of 11.81% is highly economically significant. Overall, the results strongly support the proposition that the Estimize consensus contains incremental information about the market's earnings expectations.

We expect that the incremental power of the Estimize consensus will increase as the diversity of contributors whose forecasts are included in the consensus increases. Columns 3-5 of Table 5 report the tests. Specifically, we sort all firms in our matched sample into terciles based on the diversity of the Estimize contributors. We then estimate the regression equation (4) for the low-diversity and high-diversity subsamples separately. Finally, we perform a Chow test to examine whether the coefficient on *FE_{10} Difference* is significantly different between the two subsamples. We find that the coefficient on *FE_{10} Difference* is 0.004 (*p*-value=0.00) in the low-diversity subsample and it is 0.007 (*p*-value=0.00) in the high-diversity sample. The difference in coefficients between high- and low-diversity subsamples is 0.003, which is statistically significant at the 5% level and economically meaningful. Overall, the results suggest that the Estimize consensus contains valuable information about the market's expectation that is not evident in the I/B/E/S consensus alone.

Next, we test directly whether the crowdsourced consensus is a superior measure of earnings expectations by comparing the earnings response coefficients (ERCs) for the I/B/E/S earnings surprise and Estimize earnings surprise. Following Gu and Xue (2008), we regress CAR[0,1] on earnings surprise, and month, year, and industry fixed effects separately for the I/B/E/S earnings surprise and the Estimize earnings surprise for our matched sample of firms.[24] We then use a Chow test to examine whether the ERC for the I/B/E/S surprise is smaller than the ERC for the Estimize surprise.

[24] Like Gu and Xue (2008) we use a matched sample of firm-quarters, so there is no need to control for firm-specific determinants of ERCs. Unlike in Gu and Xue (2008) sample, in our sample the median number of analysts is the same for both I/B/E/S and Estimize, so number of analysts should not have an impact on the strength of ERCs.

Panel A of Table 5 presents these results. Columns 1 and 2 show that the ERC is 1.540 for the I/B/E/S surprise and 1.912 for the Estimize surprise. The difference between the two ERCs is -0.372 (*p*-value=0.08), indicating that the Estimize earnings surprise elicits a 24% stronger market reaction. The result is consistent with the hypothesis that the Estimize consensus is a superior representation of the market's earnings expectations. Columns 3-6 examine whether the superiority of the Estimize consensus increases with greater diversity of the Estimize contributors. In the low-diversity tercile, there is no significant difference in ERCs between I/B/E/S and Estimize. In contrast, in the high-diversity tercile, the ERC for I/B/E/S is 2.197 and the ERC for Estimize is 3.974. The difference in ERCs is highly economically and statistically significant at -1.777 (*p*-value=0.00), and suggests that the market's reaction to the Estimize surprise is 81% stronger than the market's reaction to the I/B/E/S surprise. Our results suggest that researchers who study the market's response to earnings surprises will benefit significantly from using the Estimize consensus, especially in situations when the Estimize consensus is based on forecasts from a diverse set of contributors.

[Insert Table 5]

An alternative way to test whether the Estimize consensus can better explain the market's initial response to earnings surprises, is to examine the unique situations where the I/B/E/S and Estimize forecast errors generate opposing signs, thus predicting different immediate reactions. We divide the firms into subsamples of positive and negative earnings surprises with respect to the I/B/E/S consensus. Within each subsample, we compare the announcement-day abnormal returns between firms with positive and negative Estimize earnings surprises. If the Estimize consensus is a better measure of earnings expectations, then we expect the immediate market reaction to have the same sign as the Estimize surprise in the situations where the I/B/E/S and Estimize surprise disagree.

Panel B of Table 5 reports the average CAR[0,1] for the negative and positive earnings surprises with respect to the I/B/E/S consensus. Consistent with prior findings, the average CAR[0,1] is -3.11% for negative earnings surprises ($FE^{I/B/E/S} < 0$) and 1.46% for positive earnings surprises ($FE^{I/B/E/S} > 0$). Focusing on the negative surprises, CAR[0,1] is more negative, -3.53%, when the announced earnings are below both the I/B/E/S and Estimize consensus ($FE^{I/B/E/S} < 0$ & $FE^{Estimize} < 0$); and it is positive, 0.07%, when the announced earnings are below the I/B/E/S consensus but above the Estimize consensus ($FE^{I/B/E/S} < 0$ & $FE^{Estimize} > 0$). The difference of 3.60% (*p*-value=0.00) is highly economically and statistically significant.

More importantly, for positive surprises, the average CAR[0,1] is larger, 2.10%, when the announced earnings beat both the I/B/E/S and Estimize consensus ($FE^{I/B/E/S} > 0$ & $FE^{Estimize} > 0$). In contrast, the average CAR[0,1] is -0.44% when the announced earnings are above the I/B/E/S consensus but below the Estimize consensus ($FE^{I/B/E/S} > 0$ & $FE^{Estimize} < 0$). The difference in returns between the two scenarios is highly significant -2.54%. Moreover, there are 1,162, out of 4,886, cases where the earnings beat the I/B/E/S consensus but fall short of the Estimize consensus. However, there are only 176 cases out of 1,880 where the earnings fall short of the I/B/E/S consensus, but beat the Estimize consensus. Taken together, the results suggest that Estimize is a particularly useful measure of market expectations in situations when the I/B/E/S analysts "walk-down" their forecasts to beatable levels (Richardson et al., 2004).

E. Trading strategy

In this section, we examine whether our finding, that the Estimize consensus is a better representation of earnings expectations, allows us to form a profitable trading strategy.

Specifically, to the extent that Estimize better reflects the market's expectations of earnings and that Estimize and I/B/E/S consensus differ, we would expect the price prior to the earnings announcement to be more closely aligned with the Estimize consensus than with the I/B/E/S consensus. Earnings announcement can generate either a positive or a negative surprise. If the Estimize consensus is above the I/B/E/S consensus, a positive surprise will be relatively smaller and a negative surprise will be relatively larger than if the Estimize consensus is below the I/B/E/S consensus. Consequently, regardless of the sign of the surprise, when the Estimize consensus is above the I/B/E/S consensus we would expect a relatively lower return than when the Estimize consensus is below the I/B/E/S consensus.

We implement this simple trading strategy as follows. On the day prior to its earnings announcement we buy the stock if the Estimize consensus is below the I/B/E/S consensus or we short the stock if the Estimize consensus is above the I/B/E/S consensus. We hold these positions for the subsequent 10, 20, and 30 trading days (including the day of the announcement) and calculate average cumulative abnormal returns. The cumulative abnormal return, CAR, is the return on the stock in excess of the return on the size and book-to-market matching portfolio. Table 6 presents the results.

For the 10-day holding period, the average cumulative abnormal return, CAR[0, 9], is only 0.371% when the Estimize consensus is above the I/B/E/S consensus and is 0.836% when the Estimize consensus is below the I/B/E/S consensus, consistent with our expectation. The difference in the average 10-day CARs is 0.465%, which is economically meaningful and statistically significant at the 10% level. The difference in the average CARs increases with the holding period to 0.592% (*p*-value=0.04) for CAR[0, 19] and 0.604% (*p*-value=0.09) for CAR[0, 29], indicating that the most benefits of the strategy accrue within one month of the earnings announcement.

[Insert Table 6]

In the previous section, we document that the Estimize consensus is especially reflective of earnings expectations when it is formed based on forecasts from a diverse set of contributors. This finding implies that our trading strategy should be more profitable for stocks with an Estimize consensus that is constructed from a diverse set of contributors. To make the strategy implementable, we use the diversity tercile-breakpoints from the previous quarter. Consistent with our expectations, the high-diversity sample of stocks generates an average CAR[0,9] of only 0.195% when the Estimize consensus is above the I/B/E/S consensus and a significant CAR[0,9] of 1.557% when the Estimize consensus is below the I/B/E/S consensus. The difference in the average CAR[0,9] is 1.362% (*p*-value=0.07), which is highly economically significant. In contrast, the sample of low-diversity stocks produces an insignificant CAR[0,9] difference of 0.207%.

Again, we find that the difference in the CARs increases with the holding period for the high-diversity subsample. The differences are 1.721% for CAR[0,19] and 2.137% for CAR[0,29], which are statistically significant at the 5% and 1% level, respectively. These differences are also economically meaningful, suggesting an annualized excess return between 17.81% and 21.51%.[25] While the strategy results in significant gross profits, an important issue is whether it remains profitable after accounting for the trading costs. Average bid-ask spread at the close on the day prior to the earnings announcement for our sample of large stocks is 0.057% and the 99th percentile spread is 0.36%. Average commission during our sample period is below $10 per trade. These numbers suggest that the profits from implementing our strategy among high-diversity consensus

[25] Assuming 250 trading days in a year, CAR[0,19] of 1.721% translates into ((1.721/20)*250=) 21.51% per year.

stocks would still be significant even after accounting for the transaction costs for reasonably sized trades.[26]

F. Robustness Tests

We conclude the paper with a set of additional analyses to ensure the robustness of our results. For our tests of consensus forecast accuracy, we consider two alternative definitions of earnings consensus. To ensure that our results are not driven by outliers, we use the consensus based on the median of all forecasts issued within 14-days prior to the earnings announcement. To ensure that our results are not driven by exclusion of relevant information in I/B/E/S, we compare absolute forecast error from the I/B/E/S consensus based on mean of all forecasts issued within 60-days prior to the announcement, to the Estimize consensus based on the mean of all forecasts issued within 14-days prior to the announcement.[27] We repeat the analysis from Table 3 and report results in Table 7. In the first 3 columns, we show that the Estimize consensus is more accurate than the I/B/E/S consensus when we use the median consensus; and that the difference in accuracy is more pronounced when there are more Estimize contributors and when the Estimize contributors are more diverse. Similarly, in columns 4-6 we find that the accuracy results are robust to using a more inclusive I/B/E/S consensus.

[Insert Table 7]

[26] We are unable to determine the average shorting costs, but given the size and liquidity of the stocks in our sample it is unlikely that shorting costs would be prohibitive. Moreover, the profits to the strategy are primarily driven by the long leg of the strategy.

[27] The 14-day horizon prior to the announcement includes 75.77% of all earnings announcements on Estimize, but only 46.07% of all earnings announcements on I/B/E/S. The 60-day horizon, however, includes 78.42% of all earnings announcements in I/B/E/S.

Next, we examine whether the superiority of the Estimize consensus remains when we use different definitions of cumulative abnormal return. In columns 1-2 of Table 8, we define CAR[0,1] as the cumulative abnormal returns in excess of the market cumulative return in trading days [0,1] relative to the earnings announcement. In columns 3-4, we consider CAR[-10,1], which is the cumulative abnormal returns in trading days [-10,1] relative to the announcement. We use this longer horizon cumulative return because Collins and Kothari (1989) and Gu and Xue (2008) show that longer horizon produces stronger associations between abnormal returns and earnings surprises in I/B/E/S data. Our results show that superiority of the Estimize consensus is robust to these alternative definitions of cumulative abnormal returns.

[Insert Table 8]

V. Conclusions

The Internet is becoming increasingly important medium for sharing financial information. On the one hand, Da, Engelberg, and Gao (2011) show that investors use the Internet to search for (demand) financial information, and that such searches are related to retail investor trading. On the other hand, Chen et al. (2014) find that investors supply relevant financial information through popular social media websites and that such information sharing predicts future stock returns and earnings surprises. In this paper, we examine whether the crowdsourcing (supply) of earnings forecasts, through an Internet-based platform, provides any value-relevant information and whether the crowdsourced consensus is a superior measure of earnings expectations.

Our results show that in the majority of cases, the crowdsourced consensus is more accurate than the traditional sell-side consensus. Consistent with the ‘wisdom-of-crowds’ prediction, the forecast accuracy of the consensus increases with the number and, more importantly, the diversity of the contributors. In addition, we find that the crowdsourced consensus is a superior measure of the market's expectations of earnings than the sell-side consensus, as demonstrated by stronger earnings response coefficients and the sign of cumulative abnormal returns being more consistent with the sign of the crowdsourced forecast errors. Overall, our findings support the importance of the Internet as a channel for crowdsourcing financial information.

We acknowledge a potential limitation of our study. Our sample period is short and the characteristics of the Estimize community may evolve over time. However, trend evidence from Figure 2 suggests that the coverage is likely to increase, in terms of users and firms, representing an even broader portion of the market. Given the documented superiority of the crowdsourced consensus, we believe that the additional analysis of crowdsourced forecasts may be a fruitful area of future research. For example, Estimize may be a useful setting in examining the dynamics of forecasting behavior and whether there is evidence of learning, increased confidence, or information cascades (Clarke and Subramanian, 2006; Hilary and Menzley, 2006; Markov and Tamayo, 2006; Mikhail, Walther, and Willis, 1997). Ultimately, it will be interesting to see how the crowdsourcing of earnings forecasts impacts the compensatory relationship, research quality, and demand for sell-side analysts.

REFERENCES

Antweiler, W., and M.Z. Frank. "Is all that talk just noise? The information content of internet stock message boards." *The Journal of Finance,* 59 (2004), 1259-1294.

Armstrong, J.S. "Principles of forecasting: a handbook for researchers and practitioners." Kluwer Academic Publishing (2001), Dordrecht, The Netherlands.

Bagnoli, M.;M.D. Beneish; and S.G. Watts. 1999. "Whisper forecasts of quarterly earnings per share." *Journal of Accounting and Economics,* 28 (1999), 27-50.

Batchelor, V.R., and P. Dua. "Forecaster diversity and the benefits of combining forecasts." *Management Science*, 41 (1995), 68-75.

Bradshaw, M. T.;M.S. Drake;J.N. Myers; and L.A. Myers. "A re-examination of analysts' superiority over time-series forecasts of annual earnings." *Review of Accounting Studies*, 17 (2012), 944-968.

Brown, L.D. "Earnings forecasting research: Its implications for capital markets research." *International Journal of Forecasting*, 9 (1993), 295-320.

Brown, L. "*I/B/E/S Research Bibliography: Sixth Edition*." I/B/E/S International Inc., 2000. New York, NY.

Brown Jr., W.D., and G.D. Fernando."Whisper forecasts of earnings per share: Is anyone still listening?" *Journal of Business Research,* 64 (2011), 476-482.

Brown, L. D.;R.L. Hagerman;P.A. Griffin; andM.E. Zmijewski. "Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings." *Journal of Accounting and Economics,* 9 (1987), 61-87.

Chan, L.K.C;J. Karceski;and J. Lakonishok. "Analysts' Conflicts of Interest and Biases in Earnings Forecasts." *Journal of Financial and Quantitative Analysis*, 42 (2007), 893-913.

Chen, H.;P. De;Y.J. Hu; andB.H. Hwang."Wisdom of crowds: The value of stock opinions transmitted through social media." *Review of Financial Studies*, 27 (2014), 1367-1403.

Chen, S., and D. Matsumoto."Favorable versus unfavorable recommendations: the impact on analyst access to management-provided information." *Journal of Accounting Research,* 44 (2006), 657–689.

Cheng, Y.; M. Liu; and J. Qian. "Buy-side analysts, sell-side analysts, and investment decision of money managers." *Journal of Financial and Quantitative Analysis,* 41 (2006), 51-83.

Clarke, J.;A. Khorana;A. Patel; and P. Rau."The good, the bad, and the ugly? Differences in analyst behavior at investment banks, brokerages, and independent research firms." Working Paper (2008).

Clarke, J., and A. Subramanian.“Dynamic Forecasting Behavior by Analysts: Theory and Evidence.” *Journal of Financial Economics*, 40 (2006), 81–113.

Clemen, R.T.“Combining Forecasts: A Review and Annotated Bibliography.” *International Journal of Forecasting*, 5 (1989), 559-581.

Collins, D., and S.P. Kothari.. “An analysis of intertemporal and cross-sectional determinants of earnings response coefficients.” *Journal of Accounting and Economics*, 11 (1989), 143–181.

Cooper, R. A.; T.E. Day; and C.M.Lewis. “Following the leader: A study of individual analysts’ earnings forecasts.” *Journal of Financial Economics*, 61 (2001), 383-416.

Cotter, J.; I.Tuna; and P.Wysocki. “Expectations management and beatable targets: how do analysts react to explicit earnings guidance?” *Contemporary Accounting Research*, 23 (2006), 593–628.

Cowen, A.; B. Groysberg; and P.Healy. “Which types of analyst firms are more optimistic?” *Journal of Accounting and Economics*, 41 (2006), 119-146.

Da, Z.;J. Engelberg; and P. Gao. “In search of attention.” *The Journal of Finance*, 66 (2011), 1461-1499.

Das, S.R., and M.Y. Chen. “Yahoo! for Amazon: Sentiment extraction from small talk on the web.” *Management Science*, 53 (2007), 1375-1388.

Das, S.; C. Levine; andK. Sivaramakrishnan. “Earnings predictability and bias in analysts’ earnings forecasts.” *The Accounting Review*, 73 (1998), 277–294.

DeBondt, W.F., and W.P. Forbes. “Herding in analyst earnings forecasts: evidence from the United Kingdom.” *European Financial Management*, 5 (1999), 143-163.

DellaVigna, S., and J.M. Pollet.“Investor inattention and Friday earnings announcements.” *The Journal of Finance*, 64 (2009), 709-749.

Drake, M.S.;K.H. Gee; andJ. Thornock. “March market madness: The impact of value-irrelevant.” *The Accounting Review*, 84 (2014), 1639-1670.

Dugar, A., and S. Nathan.“The effect of investment banking relationships on financial analysts’ earnings forecasts and investment recommendations.” *Contemporary Accounting Research*, 12 (1995), 131-160.

Francis, J., and D. Philbrick.“Analysts’ decisions as products of a multi-task environment.” *Journal of Accounting Research*, 31 (1993), 216–230.

Giannini, R.; P. Irvine; and T. Shu.“The convergence and divergence of investors’ opinions around earnings news: Evidence from a social network.” Working Paper (2013).

Gu, Z., and J. Xue. “The superiority and disciplining role of independent analysts.” *Journal of Accounting and Economics*, 45 (2008), 289–316.

Hilary, G., and C. Hsu.“Analyst forecast consistency.” *The Journal of Finance*, 68 (2013), 271–297.

Hilary G., and L. Menzly. “Does Past Success Lead Analysts to Become Overconfident? “ *Management Science*, 52 (2006), 489 – 500.

Hill, S., and N. Ready-Campbell.“Expert stock picker: the wisdom of (experts in) crowds.” *International Journal of Electronic Commerce*, 15 (2011), 73-102.

Jacob, J.; S. Rock; and D. Weber. “Do non-investment bank analysts make better earnings forecasts?” *Journal of Accounting, Auditing and Finance*, 23 (2008), 23-61.

Jackson, A. R. “Trade generation, reputation, and sell-side analysts.” *The Journal of Finance*, 60 (2005), 673-717.

Jame, R.; R. Johnston; S. Markov; and M. Wolfe. “The Value of Crowdsourced Earnings Forecasts.” Working Paper (2015).

Ke, B., and Y. Yu, Y. “The effect of issuing biased earnings forecasts on analyst’ access to management and survival.” *Journal of Accounting Research*, 44 (2006), 965-999.

Kothari, S. P.“Capital markets research in accounting.” *Journal of Accounting and Economics*, 31 (2001), 105-231.

Lehavy, R.; F. Li; and K. Merkley. “The effect of annual report readability on analyst following and the properties of their earnings forecasts.” *The Accounting Review*, 86 (2011), 1087-1115.

Lim, T.“Rationality and analysts' forecast bias.” *The Journal of Finance*, 56 (2001), 369-385.

Malmendier, U., and D.M. Shanthikumar. “Are Small Investors Naive about Incentives?” *Journal of Financial Economics*, 85 (2007), 457–489.

Markov, S., and A. Tamayo. “Predictability in Financial Analyst Forecast Errors: Learning or Irrationality?” *Journal of Accounting Research*, 44 (2006), 725–761.

Matsumoto, D. ”Management’s incentives to avoid negative earnings surprises.” *The Accounting Review*, 77 (2002), 483–514.

Michaely, R., and K.L. Womack. “Conflict of interest and the credibility of underwriter analyst recommendations.” *Review of Financial Studies*, 12 (1999), 653-686.

Mikhail, M.B.; B.R. Walther; and R.H. Willis. “Do security analysts improve their performance with experience?” *Journal Accounting Research*, 35 (1997), 131-157.

Mikhail, M.B.; B.R. Walther; and R.H. Willis. “When Security Analysts Talk, Who Listens?” *The Accounting Review*, 82 (2007), 1227–1254.

Philbrick, D., and W. Ricks. “Using value line and IBES analyst forecasts in accounting research.” *Journal of Accounting Research*, 29 (1991), 397-417.

Ramnath, S.; S. Rock; and P. Shane. “Value Line and I/B/E/S earnings forecasts. “ *International Journal of Forecasting*, 21 (2005), 185−198.

Richardson, S.; S.H. Teoh; P.D. Wysocki. “The Walk-down to Beatable Analyst Forecasts: The Role of Equity Issuance and Insider Trading Incentives.” *Contemporary Accounting Research*, 21 (2004), 885–924.

Surowiecki, J. “The Wisdom of Crowds.” Doubleday: New York (2004).

Timmermann, A. “Forecast Combinations.” In: G. Elliot, C.W.J. Granger and A. Timmermann (eds.), Handbook of Economic Forecasting (2004). Elsevier, Amsterdam, The Netherlands.

Trueman, B. “Analyst forecasts and herding behavior.” *Review of Financial Studies*, 7 (1994), 97-124.

Tumarkin, R., and R.F. Whitelaw. “News or noise? Internet postings and stock prices.” *Financial Analysts Journal*, (2001), 41-51

Wang, G.; T. Wang; B. Wang; D. Sambasivan; Z. Zhang; H. Zheng; and B.Y. Zhao. “Crowds on Wall Street: Extracting Value from Collaborative Investing Platforms.” Working Paper (2015).

Welch, I. “Herding among security analysts.” *Journal of Financial Economics*, 58 (2000), 369-396.

APPENDIX

Variable Name	Description
#Analysts****	#Analysts is calculated as the number of analyst issuing forecasts within 14 days of the earnings announcement. Superscript **** indicates whether it is the number of Estimize or I/B/E/S analysts.
AFE**** 14	Absolute forecast error (AFE) is calculated as the absolute difference between actual earnings (from I/B/E/S) and the median forecast based on all forecasts issued within 14 days of the earnings announcement, scaled by price at the end of the quarter prior to the announcement. Superscript **** indicates whether the median forecast is obtained from the I/B/E/S forecasts or the Estimize forecasts.
B/M	Book-to-market ratio is calculated as book value of equity, the sum of shareholder equity (CEQQ) and deferred taxes and investment credit (TXDITCQ), scaled by market value, the number of shares outstanding (CSHOQ) multiplied by the quarter's end price (PRCCQ).
CAR[X,Y]	Cumulative abnormal return is calculated as the difference between the buy-and-hold return of the announcing firm and that of a size and book-to-market (B/M) matching portfolio over the window [X,Y] in the days surrounding the announcement date. Each stock is matched with 1 of 25 size and B/M portfolios from Kenneth French's website.
Dispersion	Dispersion is the standard deviation of I/B/E/S forecasts issued within 90-days of the announcement, scaled by price at the end of the quarter prior to the announcement.
Diversity	Diversity is the number of different contributor backgrounds contained in the consensus. For I/B/E/S, diversity is equal to one since all analysts are sell side. For Estimize, we count a number of different users' biographical backgrounds that are contained in the consensus. Biographical backgrounds include asset manager, broker, endowment fund, financial advisor, fund of funds, hedge fund, independent research, insurance firm, investment bank, mutual fund, pension fund, private equity, proprietary trading firm, venture capital, wealth manager and other for professionals; academia, consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, student, telecommunication services, utilities, and non-disclosed for non-professionals.
Diversity$^{\perp}$	Residual diversity is a residual from the regression of Diversity on Estimize indicator, the number of Estimize contributors, and the interaction term between the two variables.
FE**** 14	Forecast error is a difference between actual earnings (from I/B/E/S) and the mean forecast based on all forecasts issued within 14-days of the earnings announcement, scaled by price at the end of the quarter prior to the announcement. Superscript **** indicates whether the mean forecast is obtained from I/B/E/S forecasts or Estimize Forecasts.
Horizon	Horizon is the median number of days between the forecast issuance and announcement day.
Size	Size is calculated as the natural logarithm of market capitalization, price (PRC) multiplied by the number of shares outstanding (SHROUT), in the month prior to quarter's end.
Reporting Lag	Reporting Lag is the number of days from the fiscal quarter end until the announcement date.
Share Turnover	Share Turnover is calculated as the average monthly turnover (Volume/Shares Outstanding) from the previous 6 months.

Table 1
Descriptive statistics

Panel A reports mean [median] firm characteristics for all I/B/E/S firms, I/B/E/S firms with Estimize coverage, and I/B/E/S firms without Estimize coverage. Panel B reports horizon distribution for forecasts issued within 14 days prior to the announcement in I/B/E/S and Estimize and the distribution of the number of forecasts issued within 14 days prior to the earnings announcement for firms that have Estimize and I/B/E/S coverage within 14 days of the announcement. The sample contains firms that have quarterly earnings forecasts for 2012-2014 available on I/B/E/S, have share code 10 or 11, have stock price above $1, and have CRSP coverage. I/B/E/S and Estimize firms must have a forecast made within 14 days of the earnings announcement. Please see the Appendix for variable definitions. All continuous variables are winsorized at the 1% and 99% level.

Panel A: Characteristics of firms by coverage

Variable	I/B/E/S All (1)	I/B/E/S with Estimize (2)	I/B/E/S no Estimize (3)	Mean diff (3-2)	Median diff (3-2)
	Mean [Median]	Mean [Median]	Mean [Median]	T-stat (*p*-value)	K-Wallis (*p*-value)
Size	21.351	22.385	20.969	-27.290	3797.578
	[21.295]	[22.246]	[20.918]	(0.000)	(0.000)
B/M	0.598	0.446	0.657	14.420	1576.597
	[0.496]	[0.359]	[0.562]	(0.000)	(0.000)
Share Turnover	0.209	0.240	0.198	-7.010	715.394
	[0.158]	[0.184]	[0.149]	(0.000)	(0.000)
$\#Analysts^{I/B/E/S}$ 14	4.100	6.001	3.419	-10.850	2133.263
	[3.000]	[4.000]	[2.000]	(0.000)	(0.000)
Dispersion (×100)	0.280	0.128	0.340	26.180	2553.129
	[0.105]	[0.057]	[0.135]	(0.000)	(0.000)
$AFE^{I/B/E/S}$ 14 (×100)	0.504	0.208	0.614	24.940	1748.697
	[0.156]	[0.082]	[0.202]	(0.000)	(0.000)
$FE^{I/B/E/S}$ 14 (×100)	0.033	0.069	0.020	-4.070	12.765
	[0.044]	[0.044]	[0.043]	(0.001)	(0.000)
Price	35.921	52.191	29.906	-19.300	2885.578
	[27.045]	[43.410]	[22.260]	(0.000)	(0.000)
# of Obs.	27,905	7,528	20,377		

Panel B: Forecasts issued within 14 days prior to the earnings announcement

	Mean	Std. Dev	P10	P25	Median	P75	P90
I/B/E/S	6.00	5.00	1	2	4	9	13
Estimize	5.59	7.22	1	1	3	6	13
Diversity	3.53	2.93	1	1	4	5	8

Table 2
Earnings forecast accuracy

Panel A reports the univariate results for absolute forecast errors by user coverage. Panel B reports the univariate results for absolute forecast error for firms covered by Professional Estimize contributors. Panel C reports the univariate results for absolute forecast error by diversity. Panel D reports the frequency for difference in absolute forecast error between I/B/E/S and Estimize. The sample contains firms that have quarterly earnings forecasts for 2012-2014, have share code 10 or 11, have share price above $1, and have coverage in CRSP, I/B/E/S, and Estimize. Using actual earnings reported in I/B/E/S, we calculate absolute forecast error as the absolute difference between actual earnings and mean forecast based on all forecasts issued within 14-days prior to the announcement date. $AFE^{I/B/E/S}$ is the absolute forecast error using the mean forecast from I/B/E/S. $AFE^{Estimize}$ is the absolute forecast error using the mean forecast from Estimize. $AFE^{Estimize\text{-}Professional}$ is the absolute forecast error using the mean forecast from Estimize contributors who identify themselves as Professionals. $AFE^{Estimize\text{-}Non\text{-}Professional}$ is the absolute forecast error using the mean forecast from Estimize contributors who identify themselves as Non-Professionals. AFE are multiplied by 100 for expositional purposes. All continuous variables are winsorized at the 1% and 99% level. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

Panel A: Accuracy by User Coverage	N	Mean (*p*-value)	Median (*p*-value)
$AFE^{I/B/E/S}$	7,528	0.190	0.083
$AFE^{Estimize}$	7,528	0.186	0.077
$AFE^{Estimize\text{-}Professional}$	5,016	0.167	0.068
$AFE^{Estimize\text{-}Non\text{-}Professional}$	6,552	0.184	0.077
$AFE^{I/B/E/S}$ - $AFE^{Estimize}$	7,528	0.004^{**}	0.005^{***}
		(0.043)	(0.002)

Panel B: Accuracy for firms with Professional Coverage	N	Mean (*p*-value)	Median (*p*-value)
$AFE^{Estimize\text{-}Professionals}$	5,016	0.167	0.068
$AFE^{Estimize}$	5,016	0.165	0.067
$AFE^{I/B/E/S}$	5,016	0.173	0.074
$AFE^{Estimize\text{-}Professionals}$ - $AFE^{Estimize}$	5,016	0.002^{**}	0.001
		(0.015)	(0.573)
$AFE^{I/B/E/S}$ - $AFE^{Estimize}$	5,016	0.008^{***}	0.007^{***}
		(0.000)	(0.000)

Panel C: Accuracy by Diversity in Estimize	Mean (*p*-value)		Median (*p*-value)	
	High diversity N=2,179	Low diversity N=3,496	High diversity N=2,179	Low diversity N=3,496
$AFE^{Estimize}$	0.133	0.054	0.232	0.103
$AFE^{I/B/E/S}$	0.148	0.044	0.240	0.103
$AFE^{I/B/E/S}$ - $AFE^{Estimize}$	0.015^{***}	-0.010^{***}	0.008^{***}	0.000
	(0.000)	(0.012)	(0.000)	(0.981)

Panel D: Accuracy Distribution by Coverage	All Estimize	Professional	Low Diversity	High Diversity
	Percent [N]	Percent [N]	Percent [N]	Percent [N]
$AFE^{I/B/E/S}$ - $AFE^{Estimize}$>0	56.72	57.79	53.17	60.73
	[4,137]	[2,840]	[1,752]	[1,319]
$AFE^{I/B/E/S}$ - $AFE^{Estimize}$<0	43.28	42.21	46.82	39.27
	[3,157]	[2,074]	[1,543]	[853]
Binomial test (*p*-value)	(0.000)	(0.000)	(0.000)	(0.000)

Table 3
Tobit regression of absolute forecast errors: Impact of the number and diversity of contributors

This table reports the coefficient estimates from the tobit regression of absolute forecast error on the number of Estimize contributors, diversity of Estimize contributors, and controls. The sample contains firms that have quarterly earnings forecasts for 2012-2014, have share code 10 or 11, have share price above $1, and have coverage in CRSP, I/B/E/S, and Estimize. The dependent variable, absolute forecast error (AFE), is calculated as the absolute difference between actual earnings and mean forecast based on forecasts issued within 14-days prior the announcement date, scaled by price. AFE are multiplied by 100 for expositional purpose. Estimize is a binary variable equal to one if the absolute forecast error (AFE) is constructed from the Estimize consensus, and zero otherwise. $\#Analysts^{Estimize}$ is a number of Estimize contributors. Diversity is a number of unique backgrounds of contributors. $Diversity^{\perp}$ is Diversity orthogonalized with respect to Estimize and $\#Analysts^{Estimize}$. $\#Analysts^{I/B/E/S}$ is a number of sell-side analysts. Horizon is a median number of days between forecast issuance and earnings announcement. Please see the Appendix for variable definitions. All continuous variables are winsorized at the 1% and 99% level. The coefficients are multiplied by 100 for expositional purposes. Standard errors (in parentheses) are heteroskedasticity-robust and clustered by announcement date. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively.

	AFE (1)	AFE (2)	AFE (3)
Intercept	22.020***	22.528***	21.723***
	(3.869)	(3.925)	(3.922)
Estimize	-3.082***	-1.656*	-3.113***
	(0.936)	(0.949)	(0.939)
$\#Analysts^{Estimize}$	-0.247***		-0.250***
	(0.049)		(0.049)
$\#Analysts^{Estimize}\times$Estimize	-0.059**		-0.070**
	(0.028)		(0.030)
Diversity		-0.993***	
		(0.116)	
$Diversity^{\perp}$			-1.716***
			(0.238)
$\#Analysts^{I/B/E/S}$	-0.305***	-0.444***	-0.301***
	(0.075)	(0.069)	(0.076)
$\#Analysts^{I/B/E/S}\times$Estimize	0.028	0.215***	0.069
	(0.048)	(0.051)	(0.048)
Horizon	-0.265**	-0.215*	-0.266**
	(0.111)	(0.110)	(0.111)
Horizon$\times$Estimize	0.615***	0.541***	0.559**
	(0.218)	(0.218)	(0.219)
Calendar Fixed Effects	Yes	Yes	Yes
Pseudo R^2	0.033	0.034	0.036
# of Obs.	15,056	15,056	15,056
# of Clusters	625	625	625

Table 4
Market reactions to earnings surprise and incremental information

This table reports the OLS regression of immediate market reaction to earnings surprise. The sample contains firms that have quarterly earnings forecasts for 2012-2014, have share code 10 or 11, have share price above $1, and have coverage in CRSP, I/B/E/S, and Estimize. Using quarterly earnings announcement dates from I/B/E/S, we calculate the average 2-day announcement cumulative abnormal returns (*CAR*[0,1]). FE_{10} is a decile rank of earnings surprise based on all I/B/E/S forecasts issued within 14 days before the announcement. FE_{10} Difference is the difference between decile ranks of earnings surprises from Estimize forecasts and I/B/E/S forecasts (FE_{10}^{E}-FE_{10}). Columns 3 and 4 are based on the terciles of the diversity measure value. F-statistics (*p*-value) are reported in column 5. Industry fixed effects are based on the Fama-French 10 industry classification. Year and Month Fixed effects are based on the earnings announcement date. Standard errors (in parentheses) are heteroskedasticity-robust and clustered by announcement date. All continuous independent variables are winsorized at the 1% and 99% level. Please see the Appendix for variable definitions. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively.

			Low Diversity	High Diversity	F-stat (difference) (3-4)
	(1)	(2)	(3)	(4)	(5)
Intercept	-0.061***	-0.070***	-0.077***	-0.058*	
	(0.013)	(0.012)	(0.017)	(0.030)	
FE_{10}	0.016***	0.017***	0.016***	0.015***	0.010
	(0.002)	(0.002)	(0.003)	(0.005)	(0.913)
FE_{10} Difference	--	0.005***	0.004***	0.007***	4.750
		(0.001)	(0.001)	(0.001)	(0.029)
Size	0.004***	0.004***	0.004***	0.004**	
	(0.001)	(0.001)	(0.001)	(0.002)	
B/M	0.001**	0.002***	0.001	0.002	
	(0.001)	(0.001)	(0.001)	(0.002)	
Ln(1+#Analysts$^{I/B/E/S}$)	-0.002	-0.000	0.002	-0.003	
	(0.004)	(0.004)	(0.005)	(0.008)	
Share Turnover	-0.002	-0.003	-0.003	-0.001	
	(0.002)	(0.002)	(0.002)	(0.003)	
Friday	-0.008	-0.008	-0.010	0.001	
	(0.006)	(0.006)	(0.008)	(0.011)	
Reporting Lag	0.001**	0.001**	0.001***	0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	
FE x Controls	Yes	Yes	Yes	Yes	
Industry Fixed Effects	Yes	Yes	Yes	Yes	
Year Fixed Effects	Yes	Yes	Yes	Yes	
Month Fixed Effects	Yes	Yes	Yes	Yes	
R^2	0.127	0.142	0.162	0.112	
#of Obs.	7,247	7,247	3,378	2,091	

Table 5
The superiority of the earnings consensus

Panel A reports earnings response coefficients (ERCs) and difference in ERCs between the I/B/E/S and Estimize earnings surprise. Panel B reports the average market reaction (*CAR*[0,1]) by the sign of earnings surprise based on the I/B/E/S and Estimize consensus, and t-test for the difference in average returns. Sample contains firms that have quarterly earnings forecasts for 2012-2014, have share code 10 or 11, have share price above $1, and have coverage in CRSP, I/B/E/S, and Estimize. Using quarterly earnings announcement dates from I/B/E/S, we calculate the average 2-day announcement cumulative abnormal returns (*CAR*[0,1]). FE is a difference between actual earnings and mean forecast based on forecasts issued within 14-days prior the announcement date, scaled by price. Industry fixed effects are based on the Fama-French 10-industry classification. Year and Month Fixed effects are based on the earnings announcement date. Standard errors (in parentheses) are heteroskedasticity-robust and clustered by announcement date in Panel A. Returns are expressed as percentages and *p*-values are reported in parentheses in Panel B. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively.

Panel A: Comparison of Earnings Response Coefficients between I/B/E/S and Estimize

	Full Sample N=7,247		Low Diversity N=3,378		High Diversity N=2,091	
Intercept	0.015***	0.016***	0.010	0.012*	0.024***	0.025***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.009)	(0.009)
$FE^{I/B/E/S}$	1.540***		1.583***		2.197***	
	(0.280)		(0.340)		(0.457)	
$FE^{Estimize}$		1.912***		1.638***		3.974***
		(0.373)		(0.406)		(0.637)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.037	0.042	0.055	0.052	0.031	0.056
$FE^{I/B/E/S}$ - $FE^{Estimize}$	-0.372*		-0.055		-1.777***	
(*p*-value of the χ^2 statistic)	(0.084)		(0.720)		(0.001)	

Panel B: Market Reaction by Forecast Error

	N	Mean (*p*-value)		N	Mean (*p*-value)
Negative Surprise: $FE^{I/B/E/S}<0$	1,932	-3.114	Positive Surprise: $FE^{I/B/E/S}>0$	5,106	1.455
$FE^{I/B/E/S}<0$ & $FE^{Estimize}<0$	1,704	-3.524	$FE^{I/B/E/S}>0$ & $FE^{Estimize}>0$	3,724	2.099
$FE^{I/B/E/S}<0$ & $FE^{Estimize}>0$	176	0.072	$FE^{I/B/E/S}>0$ & $FE^{Estimize}<0$	1,162	-0.438
Difference:	1,880	3.596***	Difference:	4,886	-2.537***
		(0.000)			(0.000)

Table 6
Trading Strategy

This table compares the performance of a trading strategy based on the difference in the mean consensus of Estimize and I/B/E/S. The sample contains firms that have quarterly earnings forecasts for 2012-2014, have share code 10 or 11, have share price above $1, and have coverage in CRSP, I/B/E/S, and Estimize. Cumulative abnormal return, CAR, is calculated as the difference between the buy-and-hold return of the announcing firm and that of a size and book-to-market (B/M) matching portfolio over the window [0,9], [0,19], and [0,29] in the days surrounding the announcement date. We implement a long-short trading strategy by investing in firms if the Estimize consensus is below the I/B/E/S consensus, and shorting firms if the Estimize consensus above the I/B/E/S consensus. We also report average cumulative returns for the strategy when focusing only on low- and high-Diversity. We determine low- and high-diversity firms based on diversity tercile-breakpoints from the previous quarter. Standard errors are robust to heteroskedasticity and clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively.

Trading Strategy: Short when Estimize consensus>I/B/E/S consensus, Long when Estimize consensus<I/B/E/S consensus	Estimize Consensus>I/B/E/S Consensus (1)	Estimize Consensus<I/B/E/S Consensus (2)	Difference (p-value) (3)
CAR [0,9]	0.371%	0.836%	0.465%*
	N=5,439	N=1,622	(0.084)
CAR [0,9]: Low Diversity	0.490%	0.697%	0.207%
	N=2,455	N=927	(0.556)
CAR [0,9]: High Diversity	0.195%	1.557%	1.362%*
	N=1,214	N=222	(0.072)
CAR [0,19]	0.477%	1.069%	0.592%**
	N=5,438	N=1,622	(0.044)
CAR [0,19]: Low Diversity	0.550%	0.778%	0.227%
	N=2,454	N=927	(0.207)
CAR [0,19]: High Diversity	0.449%	2.170%	1.721%**
	N=1,214	N=222	(0.021)
CAR [0,29]	0.659%	1.263%	0.604%*
	N=5,436	N=1,621	(0.090)
CAR [0,29]: Low Diversity	0.809%	0.847%	0.038%
	N=2,454	N=926	(0.938)
CAR [0,29]: High Diversity	0.533%	2.670%	2.137%***
	N=1,214	N=222	(0.009)

Table 7
Robustness tests for forecast accuracy.

This table reports the tobit regression for alternative measures of absolute forecast error on the number of Estimize contributors, diversity of Estimize contributors, and controls. The sample contains firms that have quarterly earnings forecasts for 2012-2014, have share code 10 or 11, have share price above $1, and have coverage in CRSP, I/B/E/S, and Estimize. The dependent variable, absolute forecast error (AFE), is calculated as the absolute difference between actual earnings and median forecast based on forecasts issued within 14-days prior to the announcement date, scaled by price in Columns 1-3. The dependent variable, absolute forecast error (AFE), is calculated as the absolute difference between actual earnings and mean forecast based on forecasts issued within 60-days (14-day) prior to the announcement date in I/B/E/S (Estimize), scaled by price in Columns 4-6. AFE are multiplied by 100 for expositional purpose. Estimize is a binary variable equal to one if the absolute forecast error (AFE) is constructed from the Estimize consensus, and zero otherwise. $\#Analysts^{Estimize}$ is a number of Estimize contributors. Diversity is a number of unique backgrounds of contributors. $Diversity^{\perp}$ is Diversity orthogonalized with respect to Estimize, $\#Analysts^{Estimize}$, and the interaction. $\#Analysts^{I/B/E/S}$ is a number of sell-side analysts. Horizon is a median number of days between forecast issuance and earnings announcement. Please see the Appendix for variable definitions. All continuous variables are winsorized at the 1% and 99% level. Standard errors (in parentheses) are heteroskedasticity-robust and clustered by announcement date. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively.

	Median Consensus			I/B/E/S 60-day Consensus		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.221***	0.229***	0.218***	18.133***	18.861	17.820***
	(0.041)	(0.041)	(0.041)	(3.562)	(3.608)	(3.616)
Estimize	-0.036***	-0.021**	-0.037***	-0.264	0.843***	-0.360
	(0.010)	(0.010)	(0.010)	(0.891)	(0.932)	(0.897)
$\#Analysts^{Estimize}$	-0.002***		-0.002***	-0.237***		-0.241***
	(0.000)		(0.000)	(0.047)		(0.047)
$\#Analysts^{Estimize}\times$Estimize	-0.002***		-0.002***	-0.093***		-0.105***
	(0.000)		(0.000)	(0.029)		(0.030)
Diversity		-0.013***			-1.056***	
		(0.001)			(0.120)	
$Diversity^{\perp}$			-0.020***			-1.745***
			(0.003)			(0.239)
$\#Analysts^{I/B/E/S}$	-0.005***	-0.006***	-0.005***	-0.168***	-0.258***	-0.165***
	(0.001)	(0.001)	(0.001)	(0.058)	(0.053)	(0.058)
$\#Analysts^{I/B/E/S}\times$Estimize	0.002***	0.004***	0.003***	0.027	0.153***	0.057
	(0.001)	(0.001)	(0.001)	(0.035)	(0.036)	(0.035)
Horizon	-0.003***	-0.003**	-0.003***	0.099**	0.114	0.099**
	(0.001)	(0.001)	(0.001)	(0.039)	(0.039)	(0.039)
Horizon×Estimize	0.007**	0.006***	0.006***	0.251	0.212	0.195
	(0.002)	(0.002)	(0.002)	(0.182)	(0.182)	(0.183)
Calendar Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.033	0.034	0.036	0.033	0.034	0.036
# of Obs.	15,056	15,056	15,056	15,056	15,056	15,056
# of Clusters	625	625	625	625	625	625

Table 8
Robustness test for market reactions to I/B/E/S and Estimize earnings surprise

This table reports the test for difference in earnings response coefficients (ERCs) for two alternative definitions of cumulative abnormal returns. Sample contains firms that have quarterly earnings forecasts for 2012-2014, have share code 10 or 11, have share price above $1, and have coverage in CRSP, I/B/E/S, and Estimize. In columns 1-2 cumulative abnormal returns is calculated as a difference in the cumulative abnormal returns on a stock and that on the market in trading days [0,1] relative to the earnings announcement. In columns 3-4 cumulative abnormal return is calculated as a difference in the cumulative returns on a stock and that of a matching size and book-to-market portfolio in trading days [-10,1] relative to the earnings announcement. FE is a difference between actual earnings and mean forecast based on forecasts issued within 14-days prior the announcement date, scaled by price. Industry fixed effects are based on the Fama-French 10-industry classification. Year and Month Fixed effects are based on the earnings announcement date. Returns are expressed as percentages. Standard errors (in parentheses) are heteroskedasticity-robust and clustered by announcement date. *, **, and *** indicate significance at the 10%, 5%, and 1%, respectively.

	Return in Excess of the Market		Longer Horizon Return	
Intercept	0.015***	0.015***	0.021***	0.022***
	(0.005)	(0.005)	(0.006)	(0.006)
$FE^{I/B/E/S}$	1.540***		1.575***	
	(0.280)		(0.429)	
$FE^{Estimize}$		1.918***		2.060***
		(0.376)		(0.490)
Industry Fixed Effects	Yes	Yes	Yes	Yes
Month Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.034	0.039	0.027	0.033
$FE^{I/B/E/S}$ - $FE^{Estimize}$	-0.378*		-0.485	
(p-value of the χ^2 statistic)	(0.082)		(0.071)	

Figure 1: Estimize Interface.

Figure 1 shows the interface that an Estimize user will see for any given firm. Figure 1 shows the Estimize platform for Lululemon Athletics Inc for the fourth quarter of the 2014 fiscal year. For each ticker symbol, the platform shows the earnings per share, company's guidance, Wall Street's consensus, and Estimize's community consensus for current and prior quarters.

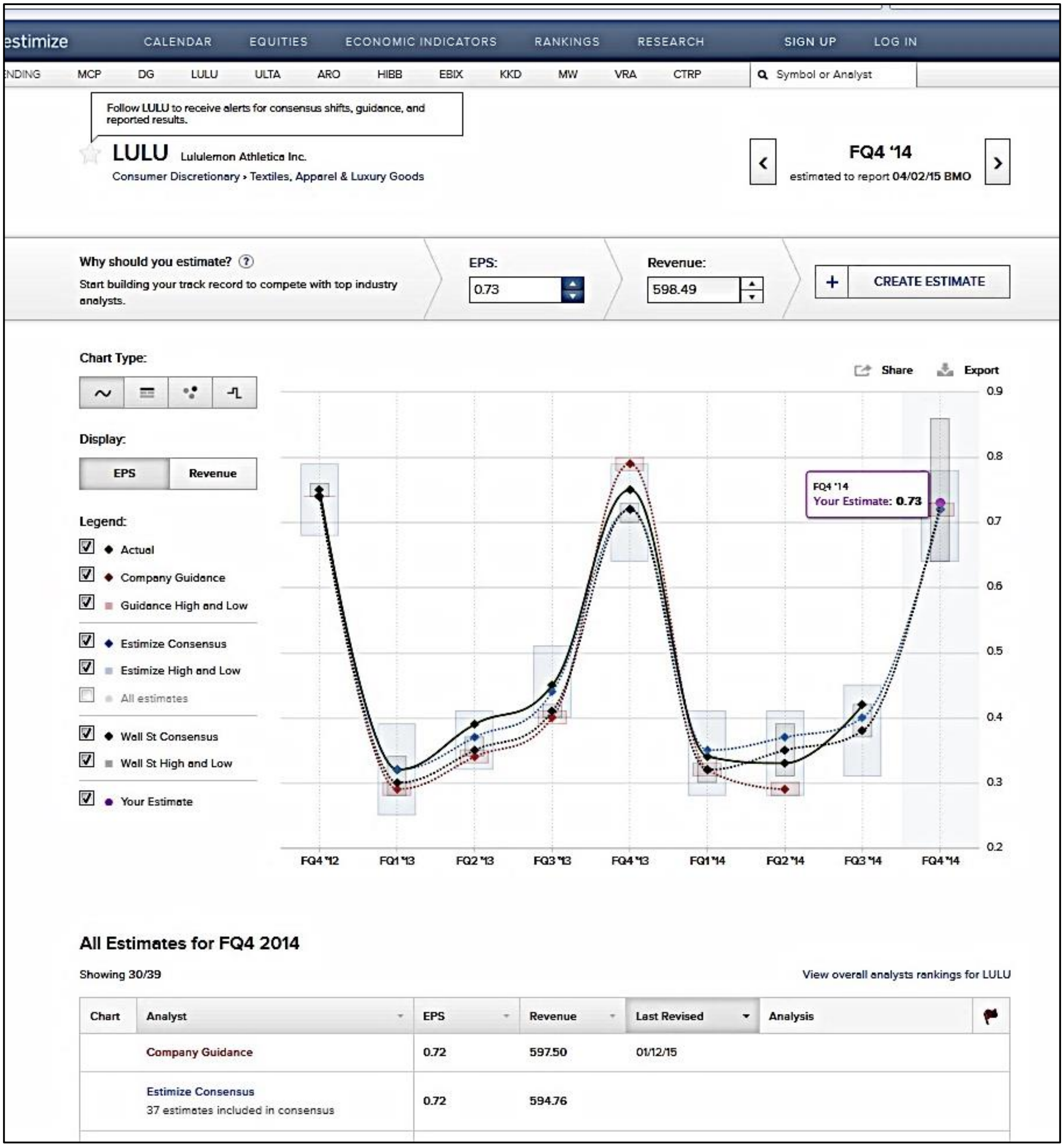
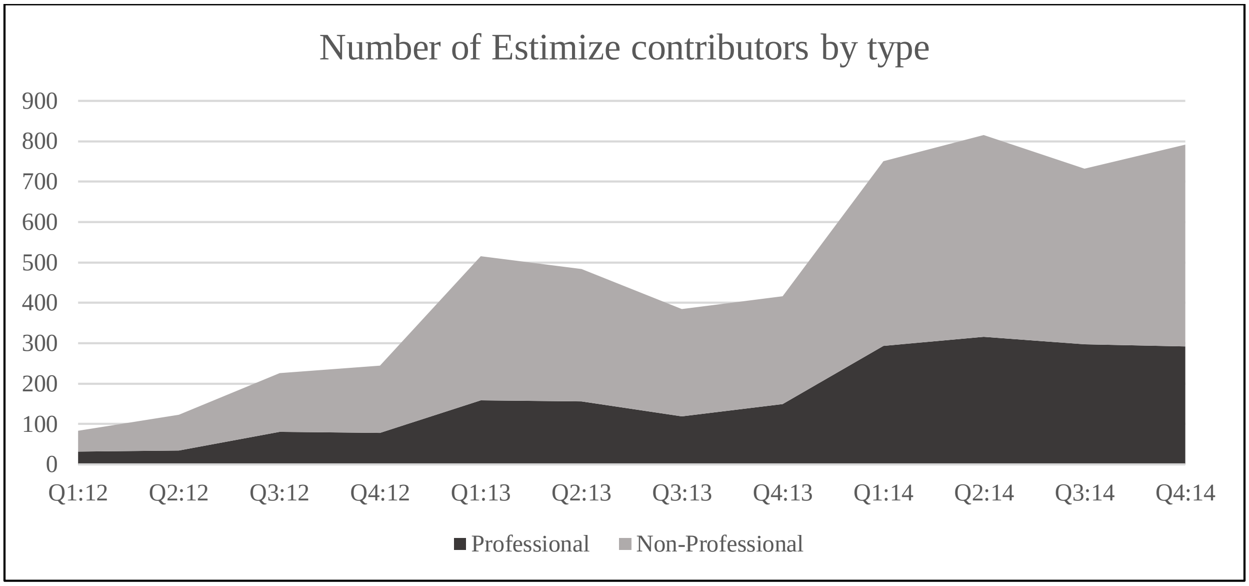


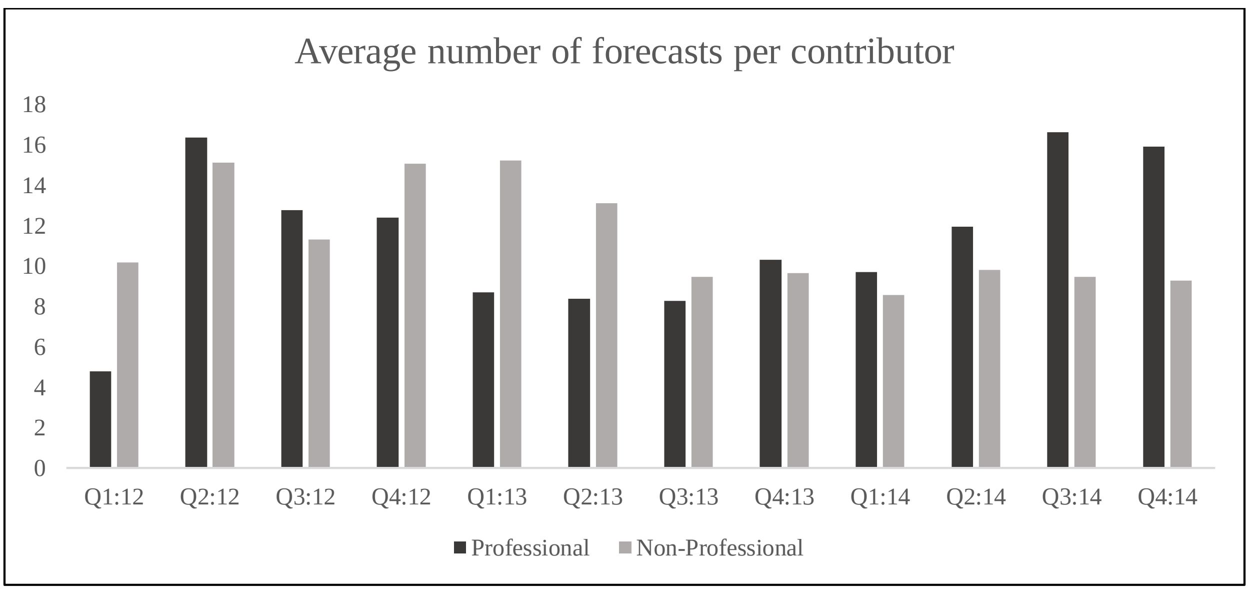
Figure 2: Time series of Estimize coverage

The Estimize sample is from Q1:2012 to Q4:2014. The sample includes only forecasts issued within 90 days prior to the earnings announcement. If a contributor issues multiple forecasts for the same firm-quarter, we only include the most recent forecast. We use contributors' reported biography to determine contributor type. X-axis shows calendar quarters. Panel A shows the total number of contributors who issued at least one earnings forecast on Estimize platform, by contributor type. Panel B shows the average number of forecasts that contributors issue on Estimize per quarter. Panel C shows total number of firms with Estimize following in the 90 (14) days prior to earnings announcement by quarter. Panel D shows average number of forecasts per firm issued on Estimize by contributor type.

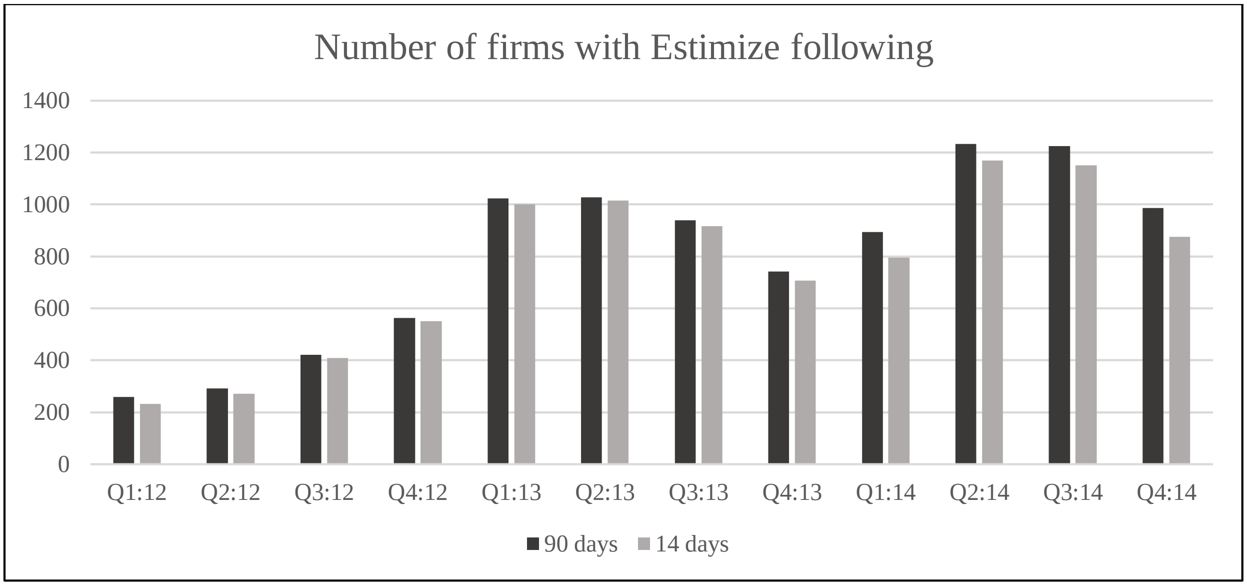
A



B

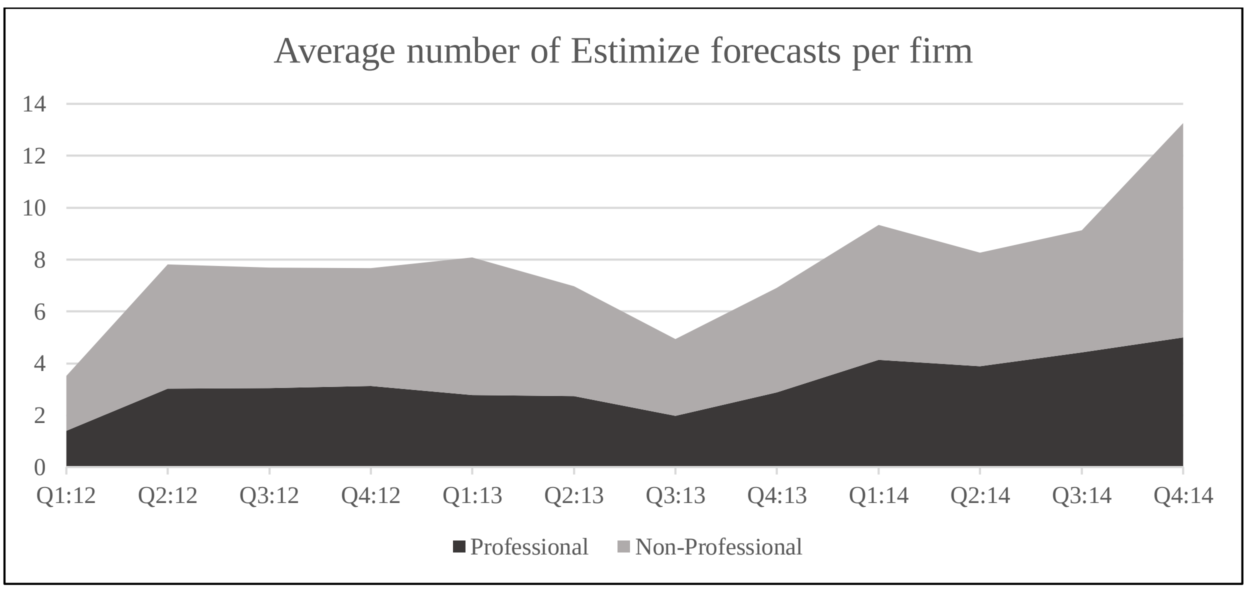


C



Number of firms with Estimize following
1400
1200
1000
800
600
400
200
0
Q1:12
Q2:12
Q3:12
Q4:12
Q1:13
Q2:13
Q3:13
Q4:13
Q1:14
Q2:14
Q3:14
Q4:14
90 days
14 days

D



Average number of Estimize forecasts per firm
14
12
10
8
6
4
2
0
Q1:12
Q2:12
Q3:12
Q4:12
Q1:13
Q2:13
Q3:13
Q4:13
Q1:14
Q2:14
Q3:14
Q4:14
Professional
Non-Professional

Figure 3: Distribution of Signed Forecast Errors

The Estimize sample is from Q1:2012 to Q4:2014. The sample includes forecasts issued within 14 days prior to the earnings announcement. If a contributor issues multiple forecasts for the same firm-quarter, we only include the most recent forecast. Forecast error (FE) is the difference between actual earnings and mean forecast, scaled by price. Positive forecast error indicates pessimism. The figure reports proportion of negative, zero, and positive forecast errors for Estimize and I/B/E/S split on their consensus accuracy (absolute forecast errors, AFE).